NASA SP–179

THE BOOK OF MARS

by
Samuel Glasstone

Scientific and Technical Information Division
OFFICE OF TECHNOLOGY UTILIZATION 1968
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C.

Photograph of Mars obtained on August 24, 1956 (18 days before the opposition on September 11, 1956) by R. B. Leighton of the California Institute of Technology. The distance between Earth and Mars at the time the photograph was taken was about 35,000,000 miles. The Mt. Wilson 60-inch reflector was used with its aperture cut to 21 inches by an off-axis diaphragm. The exposure time, on Kodachrome Type A film, was 20 seconds. The positive, used in making the print, was composed by George Emmerson at the Jet Propulsion Laboratory.

This color photograph suggests that the darker areas of Mars are not necessarily "green" in color as they are often described, but may be a darker shade of the prevailing yellow-orange light areas. It is noted that the photograph as it appears here has been subjected to duplication in the course of which some minor color changes occurred. The brilliant white south polar cap is clearly evident. Rather surprisingly, this cap is probably just what it looks like—a thin layer of frozen water, perhaps in the form of hoarfrost. As the polar cap recedes, the dark areas (especially those in the same hemisphere) become darker. The dark area near the lower right-hand limb of Mars is Syrtis Major, one of the most prominent and well-known features of the planet. This feature, among others of its kind, has been of increasing interest to exobiologists in recent years. The extremely light-colored area to the right and just below the ice cap is Hellas, one of the most prominent Martian desert areas.

Foreword

TODAY WHEN KNOWLEDGE GROWS at an ever-increasing rate, we are becoming acutely aware of man's influence, as well as his dependence, on his environment. Problems of pollution and depletion of resources arise from man's use and misuse of his environment. Knowledge and understanding, as well as technological capability and social determination, will be required to solve these global difficulties.

In a very real sense, the solar system is man's environment. The Sun furnishes life-giving energy to our planet and literally controls the behavior of Earth's atmosphere. The Moon and planets collectively hold the key to that full understanding of planetary origin and evolution that we will need in order to solve most effectively our challenges on Earth. Through study of both their similarities to and differences from the Earth, the planets will provide greater insight into the mysteries of our own planet.

Mars and Venus, our closest planetary neighbors in space, are similar to Earth in many respects, although different in many others. Mars in particular has long attracted the attention of the astronomer. With its transparent atmosphere, the surface has been sufficiently visible to show features that come and go with the seasons. Speculation that the variable markings may indicate the presence of some form of life has enhanced interest in the red planet. Long before the advent of modern astronomy, surely even before the dawn of history, Mars was known to the peoples of the world. We find the planet identified by ancient civilizations and see it appearing in Western and other mythologies. Except for the Sun and Moon, it has probably longer been a subject of wonder and awe to man than any other celestial object.

Over the past century and a half, Mars came to be known as one of the most Earth-like objects in the solar system, possessing a thin atmosphere and exhibiting white polar caps that seasonally extend and retract, and dark markings that come and go with seasons. With a rotation period comparable to that of the Earth, it appears to have a rudimentary weather system, involving fierce winds and occasional duststorms of very broad extent.

There came a time when telescopes and their associated instrumentation had revealed most of what they could tell about the planets. This factor, together with growing interest in the exciting and fundamental field of stellar and galactic astronomy,

led to a period of marked decline in astronomical interest in the planets. This interest was renewed as a result of our increased capability to observe and measure, which came about in two ways: through improved instrumentation for Earth-based observing, and through man's dramatic new ability to send instrumented spacecraft to the planets themselves. The latter is illustrated by the historic encounter with Mars of the spacecraft Mariner IV, which radioed 150 million miles home to Earth its television pictures of a rugged, cratered surface reminiscent of that of the Moon, and which made the first magnetospheric, ionospheric, and atmospheric measurements from the vicinity of the planet. In the light of such new scientific potential, the Space Science Board of the National Academy of Sciences has recommended high priority for planetary exploration, with primary initial emphasis on Mars, seeking evidence of life elsewhere than on Earth, and seeking fundamental knowledge of the origin and evolution of planets in the solar system.

This book on Mars, by Dr. Samuel Glasstone, one of the foremost scientific authors of today, is published by NASA in support of the growing interest in the study of our solar system. The book is a one-volume compaction of information about Mars, gained over many years by many scientists throughout the world, using increasingly powerful and sensitive instruments, gifted insights, and rigorous induction to produce a sizable body of knowledge and theory about the red planet. As extensive as our knowledge and speculation about Mars may be today, it may turn out in the not too distant future to have been only preliminary, as new techniques of Earth and space observation help to pry open the secrets of our solar system, planet by planet.

HOMER E. NEWELL
Associate Administrator
National Aeronautics and Space Administration

Contents

I Introduction to Mars 1

II Mars: Historical Background 7

III Mars in the Solar System 33

IV Mars as a Planet 57

V The Atmosphere of Mars 73

VI The Surface of Mars 99

VII The Clouds and Haze of Mars 137

VIII The Origin of Life 153

IX The Formation of Prebiological Materials 177

X The Possibility of Life on Mars 209

XI The Detection of Life on Mars 225

XII Avoidance of Contamination of Mars 247

XIII Exploration of Mars With Spacecraft 273

I

Introduction to Mars

MARS, the fourth planet from the Sun, revolves in an orbit lying between those of Earth, the third, and Jupiter, the fifth planet. From prehistoric times Mars has attracted interest; first, because of its unusual reddish color, then because of the difficulty of understanding its apparent motion in the sky, and later, after the invention of the telescope, because it is the only planet on which surface details and the changes they undergo can be distinguished.

The planet Mars has already played an important part in making a revolutionary contribution to knowledge, and it may well do so again during the 20th century. As will be seen in the next chapter, it was from a detailed and prolonged study of the motion of Mars that the German astronomer Johannes Kepler discovered, in the early 17th century, the basis of modern views on planetary orbits. This event was of outstanding significance because, once and for all, it resulted in the overthrow of the idea, which had dominated man's thinking for many centuries, that Earth was the center of the universe (fig. 1.1).

As for the future, Mars will undoubtedly be the first planet—other than Earth—upon whose surface instruments will be landed to transmit scientific information and pictures back to Earth for extended periods of time. Venus is usually the nearest planet to Earth, and instrumented spacecraft, such as the Soviet Venera 4 in October 1967, may have already reached that planet. But the high temperatures at the surface of Venus would soon render both instruments and radio transmitters inoperative. Furthermore, the continued exploration of space will inevitably lead to the landing of men on Mars, rather than on Venus. In addition to the high surface temperature, the high atmospheric pressure and perpetual cloud cover would make the latter planet an extremely difficult objective for manned exploration.

The temperatures on Mars, on the other hand, are not greatly different from those on Earth. Except for very occasional dust-storms, the visibility is expected to be good. It is true that Martian gravity is less than that of Earth, but it is more than twice as great as the gravity on the Moon. The days on Mars are of the familiar length, about 24 hours, and there are four seasons somewhat resembling those on Earth, but almost twice

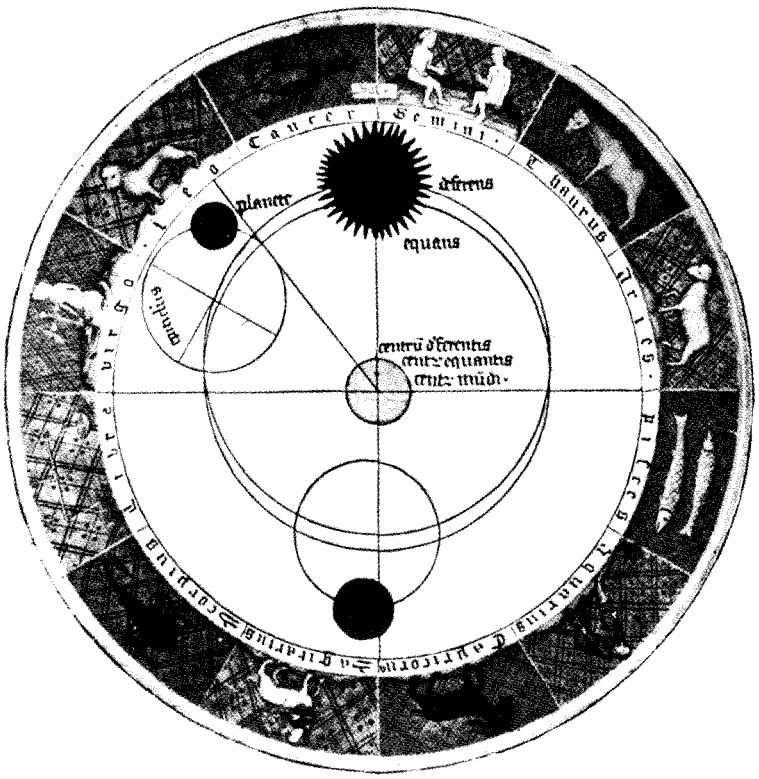


FIGURE 1.1. Earth at the center of the universe. (From 15th Century manuscript in Pierpont Morgan Library; courtesy Time, Inc.)

as long. There are a few important differences between the two planets, such as the very low atmospheric pressure on Mars, the lack of oxygen, and the absence of bodies of water, but these would not make the manned exploration of Mars any more difficult than that of the Moon.

THE POSSIBILITY OF LIFE ON MARS

The similarities between Earth and Mars have led to speculations that living creatures may exist on the latter planet. Toward the end of the 19th century some writers suggested that the strange linear features on the Martian surface, the so-called canals, were an indication of an advanced civilization. Most scholars could not accept this view, but "men from Mars" became common in science fiction.

In view of the almost complete lack of oxygen and the absence of any significant quantity of liquid water, it is highly improbable that there are now any advanced life forms on Mars. Nevertheless, as will be seen in the course of this book, it is not at all impossible that living organisms, specially adapted to the conditions on the planet, can exist or could have existed in the past. The discovery of life, past or present, on Mars would constitute one of the most exciting scientific events of all time. For this reason alone, the exploration of Mars, by instruments and by

man, can be well justified. It would be particularly appropriate if a study of Mars, the planet which dethroned Earth from its position as the center of the universe, should also prove that Earth is not the only abode of life in the solar system.

The first attempts to discover life on Mars will be made by landing instruments designed to detect phenomena characteristic of life on Earth. The results may be negative because living organisms may have developed quite differently in the Martian environment. Thus the question of whether or not life exists on Mars may not be answered until man sets foot on this planet and brings back samples of the soil for detailed study on Earth. Even if these studies are not conclusive, they may nevertheless reveal the presence of certain chemicals believed to be the precursors of life. The identification of such prebiological materials on Mars would be a discovery of outstanding significance.

THE PROBLEMS OF MARS

The search for evidence of life or of life-related substances is of the greatest general interest, but the planet Mars presents a number of problems which have intrigued scientists for many years. The answers would not only clarify the situation concerning Mars, but they would also help in the understanding of many matters related to Earth. In order to explain some of the problems of Mars, a brief outline will be given here of certain aspects of the planet. The various subjects mentioned are treated more fully in later chapters of this book.

Observations of Mars through the telescope have revealed three distinct types of surface: namely, polar caps, bright areas, and dark areas. The brilliant white polar caps are formed in the region of each Martian pole during the local autumn and winter. The caps gradually diminish in size in the spring and summer and start to form again in the autumn in each hemisphere in turn. Some scientists argue that the polar caps consist of solidified water, similar to hoarfrost, whereas others think they are mainly solid carbon dioxide. It is true that water frost is more readily deposited but, on the other hand, the Martian atmosphere contains much more carbon dioxide than it does water vapor. The first problem is then: What is the composition of the polar caps?

The bright areas, covering much of the surface, are generally reddish or orange in color. It is to these areas that the planet owes its characteristic appearance as seen by the naked eye. The great majority of the so-called canals, which are usually darker in color, cross the bright areas. The large dark regions occupy about one-fourth of the Martian surface and are mainly in the northern hemisphere. The dark areas and the canals are much lighter in color during the local winter; they become darker in the course of the spring and summer in each hemisphere. Some dark areas appear to remain constant in size, whereas others are quite variable. Occasionally a new dark area seems to develop and it may last from a few months to many years. Conversely, an existing dark area may fade out and perhaps disappear.

Most students of Mars would probably agree that the bright areas have a covering, at least, of a reddish-colored mineral containing iron. But what is the composition of the material on the dark areas? Why do these areas show seasonal changes? Are these changes caused by vegetation, as some have suggested, or is there another explanation? Why are some dark areas permanent whereas others are variable? What are the so-called canals?

It has long been assumed that, as on the Moon, the dark areas on Mars are lowlands,

whereas the bright areas are uplands. But is this necessarily true? There are arguments which suggest that the reverse may be the case. It would be of great interest to know more about the relief of the Martian surface. For example, are there really elevation differences of as much as 10 to 20 kilometers (32 800 to 65 600 feet), as has been suggested? If such elevations occur in the equatorial region, one of the strange problems concerning the flattening of Mars in the polar direction would be solved.

Although Mars has an atmosphere, it is now known to be a very tenuous one. The atmospheric pressure at the surface of the planet is roughly one-fiftieth to one-hundredth part, that is, about 1 or 2 percent, of the atmospheric pressure on Earth. The Martian atmosphere apparently consists mainly (or contains a considerable proportion) of carbon dioxide. It is expected that nitrogen and argon will also be present, but are the amounts as small as some have suggested? An answer to this question might help to explain the origin of the terrestrial atmosphere.

The temperature variations with increasing altitude in Earth's atmosphere are quite complex and not completely understood. The situation is complicated by the presence of large bodies of water as well as by the constituents of the atmosphere. Mars has no seas or lakes and its atmosphere is much less dense than that of Earth. A knowledge of the temperatures in the Martian atmosphere would undoubtedly clarify some of the problems on Earth. Terrestrial meteorologists would also benefit from information on the related problem of the wind structure on Mars and its variation with the seasons.

Clouds and haze of various types frequently appear in the atmosphere of Mars. White (or blue-white) clouds of different types are fairly common, but they usually cover only relatively small areas. Are these clouds droplets of water or small ice crystals? Or are they small particles of solid carbon dioxide? Occasionally yellow clouds form and cover very large areas of the planet. Presumably these clouds consist of dust particles of the same material as covers the bright areas on the surface. But other ideas have been suggested, and they need to be studied further.

A baffling Martian phenomenon is the so-called "blue haze" or "violet layer." If Mars is viewed (or photographed) through a telescope with a red or orange filter, or without a filter, the details of the surface are generally clearly visible. But with a blue filter, which passes only blue light, the details disappear, although the polar caps are seen to remain bright. Occasionally, there is a local or planetwide clearing for periods of hours to a few days, and then a photograph in blue light will show the surface details.

What is responsible for these strange effects? Most explanations, none of which is really satisfactory, attribute the phenomena to something in the atmosphere. They may, however, be caused largely by certain properties of the Martian surface. More must be known about the light and dark areas, as well as about the atmosphere, before the mystery of the blue haze and its occasional clearing can be solved.

Finally, what about the interior of Mars? It can be surmised, from the low density of the planet, compared with that of Earth, and the essential absence of a magnetic field, that the interiors of Mars and Earth differ in some respects. Earth probably has a core consisting mainly of the heavy metals, iron and nickel, partly molten and partly solid. If Mars has such a core at all, it is expected to be small. It would be of considerable interest to know if Mars has a heavy-metal core. This is related to the question of whether or not there has been melting in the interior of

the planet, such as is believed to have occurred in Earth. Information about the interior of Mars would undoubtedly throw light on the origin of the solar system and on the early history of Earth.

DIFFICULTIES IN OBSERVING MARS

Mars is a highly interesting subject for study, but it is also a difficult one. Both Earth and Mars orbit about the Sun, but Earth travels at a faster speed in a smaller orbit. Consequently, the distance between Earth and Mars is always changing. At intervals of about 26 months, however, the two planets come fairly close to each other, and these occasions, which are called apparitions by astronomers, provide the opportunity for studying Mars at moderately close range. The opportunities are not all equally good because the distance between Mars and Earth, even at closest approach, is not always the same.

At best, Mars can be observed for a period of a few months every 2 years or so. On the most favorable occasions, when the planet is closest to Earth, it is best observed from Earth's Southern Hemisphere. But the majority of astronomers and the most suitable telescopes are in the Northern Hemisphere. Another restriction is that at certain apparitions the north pole of Mars is directed toward Earth and then the south polar regions cannot be seen; in other years the situation is reversed. At some apparitions, however, essentially the whole surface of the planet is visible.

The rotation of Mars about its axis introduces another problem because there is a continuous change in the image seen in the telescope. The American astronomer R. S. Richardson, who has spent many years observing Mars, says that a sketch should be completed "in no more than 15 minutes," otherwise the surface features are "scrambled up" owing to the distortion of distances. It is true that, as a result of the rotation, the whole surface of Mars can be seen. But because the planet is visible for only a few hours at a time, a limited portion can be viewed each night.

Furthermore, because Mars takes but a little (37 minutes) over 24 hours to make a complete rotation, the parts of the planet that are visible to an observer on Earth do not change very much from one night to the next. As a general rule, about 5 to 6 weeks are required for the complete circumference of Mars to be seen. During this period, the distance between Earth and the planet will have changed significantly, and so also will its apparent size.

It should be borne in mind, too, that Mars is, at best, a "difficult" astronomical object. For example, in 1906, the Harvard astronomer William H. Pickering wrote: "To understand how the planet appears through a large telescope, we may examine the Moon some night through a small opera glass. The sharpness and amount of detail visible in the two cases will be similar. . . ."

Up to a point, the ability to discern details of the Martian surface can be improved by an increase in the aperture of the telescope, up to about 36 inches. The useful limit in this respect is determined by the condition of Earth's atmosphere. This is perhaps the most important aspect of the telescopic observation of Mars. "If the atmosphere is disturbed . . . " writes R. S. Richardson, "the image of Mars in even the best telescope will be blurred and shaky. . . . " The occasions of what the astronomers call good "seeing" are rare and unpredictable, and it is only when the seeing is good that there is any possibility of observing details of the Martian surface. Observatories are frequently located at high altitudes to improve seeing conditions.

Because of the limitation set by seeing, telescopes with very large apertures do not show any greater detail of the Martian surface than do smaller instruments. According to R. S. Richardson, the planet "looks as if all the color has been washed out of it" when viewed with the full aperture of the 100-inch reflector telescope of the Mount Wilson Observatory, California. "The sharpness of the image," he goes on to say, "can be considerably improved by diaphragming down, or partially covering, the secondary mirror."

Although telescopic photographs of Mars are commonly taken for purposes of record, it is recognized by astronomers that the eye can detect more detail than can be captured on film. This is one reason that photographs do not show any obvious linear features like the canals. Nevertheless, photographs are of great value in providing accurate locations of specific areas; the details can then be partially filled in by visual observation. In this way, fairly accurate maps have been prepared showing the principal features of the surface of Mars. A recent example of such a map is given in figure 2.21 at the end of chapter II.

Besides the telescope, other instruments have been used to obtain information concerning Mars. Since about the middle of the 19th century, the spectroscope has been added to the telescope to analyze the spectrum of the Martian atmosphere. As a consequence of the great distance of the planet, however, even at closest approach, and because the light from Mars must pass through the terrestrial atmosphere, most of the results have been unreliable. It is only since about 1949 that any useful data have been obtained. Other optical techniques, such as a study of the polarization and spectrum of the light reflected from the Martian surface, have also been used, although the information they provide is limited. In recent years, radar methods have been applied in the study of Mars, but the interpretation of the results may be open to question.

SPACECRAFT FOR THE OBSERVATION OF MARS

On the whole, it appears that the telescope, the spectroscope, and radar cannot alone provide the answers to many questions raised here and in other chapters of this book. What then can be done to provide more information about Mars? The solution to the problem is to make use of spacecraft which can carry instruments to the vicinity of the planet and, in due course, drop them onto its surface. The kinds of measurements that can be made with such instruments are described in chapter 13.

Ultimately, men will travel to Mars and land there. In 1965, the Mariner IV space vehicle passed within 9846 kilometers (6118 miles) of the surface of the planet and transmitted information concerning the Martian atmosphere, its surface, and its magnetic field. This is only a beginning of what can be a new era in the study of the solar system and even of the origin of life itself.

FOR FURTHER INFORMATION

A few selected references to the literature, mostly recent, are given at the end of each of the subsquent chapters. They have been chosen, in general, for their intrinsic interest and because they provide further references to the subject material of the chapter.

II

Mars: Historical Background

THE MOTIONS OF THE PLANETS

The Planets of the Ancients

FOUR THOUSAND YEARS AGO, soothsayers, watching the skies for omens from which they might predict the future, observed that the stars circled the heavens every night from east to west. With the exception of a mere handful, however, the stars always seemed to remain in fixed positions with respect to one another. As an aid to their identification, these so-called fixed stars were later divided into groups now known as constellations (fig. 2.1). Within each of the constellations the stars formed an apparently unchanging pattern, and every constellation always retained the same position in the sky in relation to the others.

In addition to the fixed stars, the ancient observers noted that, apart from the Sun and the Moon, there were five celestial bodies which looked like stars but behaved in a different manner. In the first place, these apparent stars did not belong to any of the constellations. Although they traversed the sky from east to west each night like the fixed stars, they were found to move, in the course of time, from one constellation to another, generally from west to east.

At certain intervals, however, these wandering stars, which were called planets (from the Greek word *planetes,* meaning wanderer), seemed to reverse their direction of motion through the constellations. Instead of traveling from west to east through the fixed stars, they would appear to move from east to west for a few weeks before resuming their normal motions. Another observed peculiarity of the planets was that their brightness varied quite considerably during the course of a year and from year to year. The brightness of the fixed stars, on the other hand, remained the same.

The seemingly erratic behavior of the planets aroused considerable interest. First, there were the astrologers, those early students of the stars who claimed the ability to foretell human events from the locations of the planets in relation to one another and to fixed stars. Then came the astronomers who attempted to develop a scientific explanation for the motions of the planets.

For many hundreds of years, in fact until

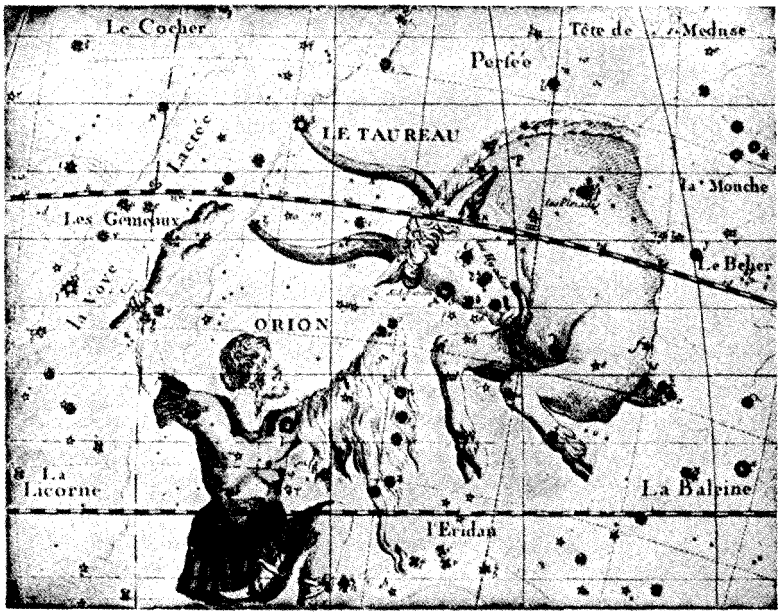


FIGURE 2.1. The constellations Orion and Taurus. (From Baker's Astronomy; courtesy D. Van Nostrand, Inc.)

the middle of the 16th century, seven heavenly bodies were recognized as planets. They were the Sun, the Moon, Mars, Mercury, Jupiter, Venus, and Saturn (fig. 2.2). The Sun and the Moon were included among the planets because they appeared, like the others, to move from west to east among the constellations. Earth was not thought of as being a planet, as it is now, because it was supposed to be the stationary center of the universe.

A reminder of the seven planets of the ancients is to be found in the names of the days of the week: Sunday (Sun); Monday (Moon); Tuesday (French, *mardi,* Mars); Wednesday (*mercredi,* Mercury); Thursday (*jeudi,* Jove or Jupiter); Friday (*vendredi,* Venus); and Saturday (Saturn).

Of the planets, one in particular attracted special attention because of its reddish color, suggestive of blood and fire. To the Sumerians and other early civilizations, this planet became a symbol of the carnage and destruction of war, and the Chaldeans, about 3000 years ago, named it Nergal for the master of battles. Later, the Greeks referred to the red planet as Ares, for their god of war, and the Romans adopted the equivalent name Mars, who was identified with Ares. In the course of time, the red planet came to be called Mars (or its equivalent) in most languages. Nevertheless, the Greek Ares still persists in such words as areocentric and areography. It may be noted that the symbol for Mars, ♂, represents a shield and spear, the ancient implements of war.

The earliest attempts to explain the apparent motions of the stars and the planets were made by the Greeks around 600 B.C. They thought that Earth was stationary, as indeed it seemed to be, and that all the heavenly bodies, including the Sun and the Moon, revolved about Earth once every day. This

FIGURE 2.2. The seven planets of the ancients. (From 17th Century engraving; courtesy Time, Inc.)

geocentric (or Earth-centered) picture of the universe seemed to agree with what the eyes saw when an observer looked at the sky, and there was no obvious reason for rejecting it. The Greek scholar Aristarchus of Samos, who lived in the third century B.C., is said to have suggested that Earth and the planets might be revolving about the Sun. But this view was not regarded with favor because of the strongly entrenched opinion among the Greeks that Earth was the fixed center of the universe.

The Ptolemaic Planetary System

A more detailed development of the geocentric concept was made in Greece during the second century B.C. by Hipparchus, the first systematic astronomer of whom there are any records. Hipparchus is said to have taken his ideas from an earlier philosopher, Apollonius of Perga (in Asia Minor), and they were adopted in turn by the Greco-Egyptian Claudius Ptolemaeus (Ptolemy), who lived in Alexandria in the early decades of the second century A.D. In his famous work that became known as the Almagest, an Arabic distortion of a word meaning "the greatest [book]," the Ptolemaic system was expounded. So convincing was Ptolemy, that his system of planetary motion remained virtually unchallenged for more than 1400 years.

According to the Ptolemaic system, Earth was stationary and around it revolved the seven planets in the following order: the Moon, Mercury, Venus, the Sun, Mars, Jupiter, and Saturn (fig. 2.3). The planets were supposed to move in a series of circles, known as deferents, with Earth at the center. In order to account for the reversal in their direction of travel through the constellations, which occurs at regular intervals, an additional motion was introduced. The planets, other than the Sun and Moon, were assumed to move on small circles, called epicycles; the centers of these epicycles were supposed to travel along the deferents. By combining the two types of motion, a qualitative interpretation could be provided for the periodic reverse (or retrograde) motion of the planets. The Sun and the Moon do not exhibit the retrograde behavior of the other planets and so they did not need epicycles.

The Greek astronomers were aware that, as seen from Earth, the planets Mercury and Venus always appeared to be close, in angular distance, to the Sun, whereas the angular dis-

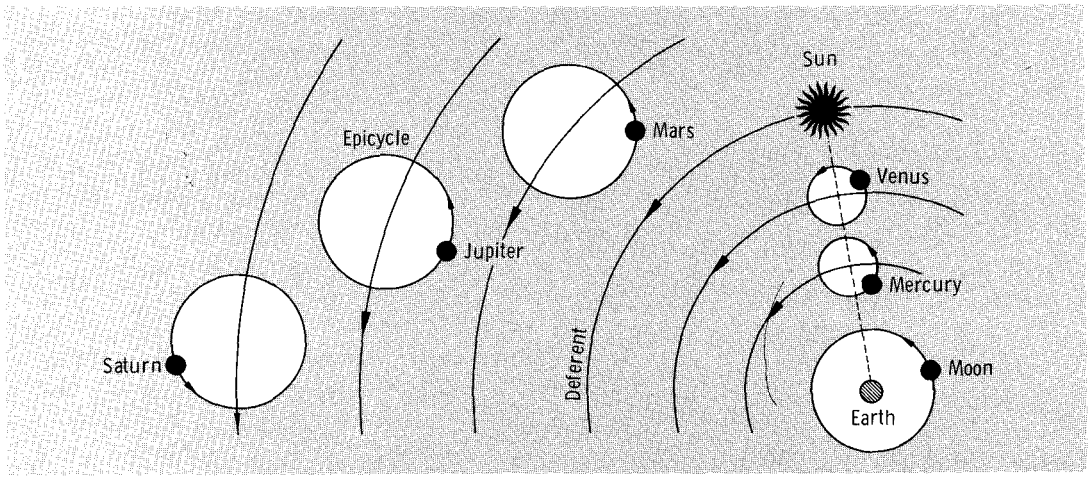


FIGURE 2.3. The Ptolemaic planetary system.

tances of the other planets—Mars, Jupiter, and Saturn—varied considerably. Consequently, the Ptolemaic system required that the centers of the epicycles of Mercury and Venus should always be on the line joining Earth and the Sun, as shown in figure 2.3.

Although the combinations of epicycles and deferents were able to provide a general picture of the apparent motions of the planets, it was not adequate to account for their actual positions over the years. Consequently, smaller, secondary epicycles were introduced with their centers moving on the primary epicycles. Eventually, more than 50 epicycles were required to represent the paths of the five starlike planets in the sky.

The Copernican Planetary System

In 1543, after more than 30 years of preparation, the Polish astronomer Nicolaus Copernicus published his great work "On the Revolutions of the Celestial Bodies," in which he revived the heliocentric (Sun-centered) concept of Aristarchus. In the Copernican system, the Sun was stationary and the planets, including Earth but not the Moon, revolved around it (fig. 2.4). Instead of assuming that the heavenly bodies circle about Earth, Copernicus showed that the nightly motion of the stars from east to west could be explained equally well by postulating that Earth rotated about its north-south axis once every day in a counterclockwise direction; that is, from west to east. The main orbit (deferent) of each planet was still taken to be circular in the Copernican system, and epicycles were necessary to account for the apparent motion of the planets.

The ideas of Copernicus appeared to offer no striking advantage or simplification over the geocentric system of Ptolemy. Furthermore, in the 16th century, belief in Earth as the center of the universe had become a religious dogma that few dared to oppose. Consequently, the Copernican system received little support. One of the arguments against it was that, if Earth did indeed move, then the directions of the nearer fixed stars, at least, should appear to change in the course of time (fig. 2.5). But no such changes could be detected. Actually, the argument was sound

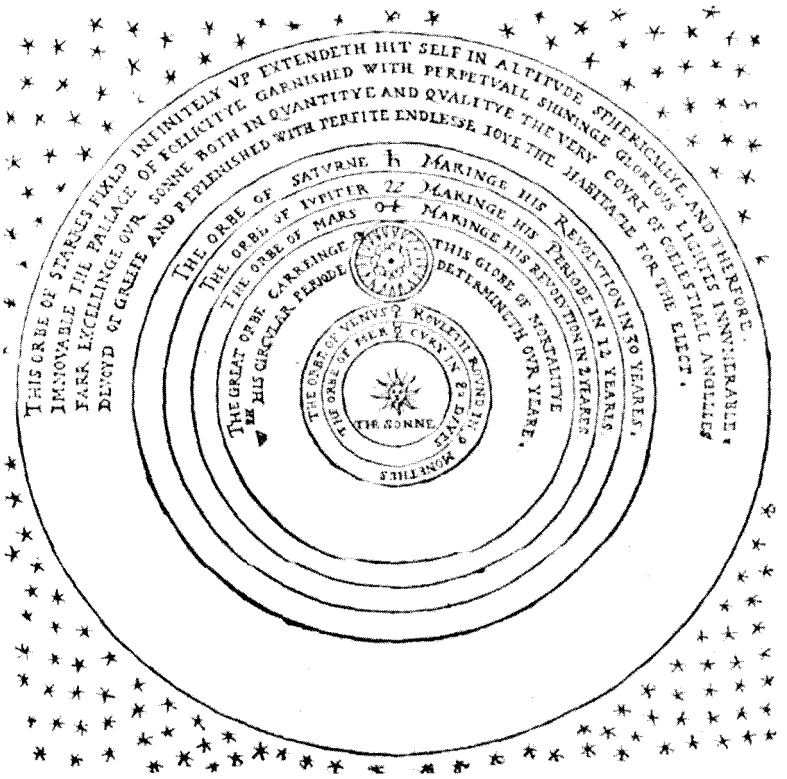

THIS ORBE OF STARES FIXED INFINITELY VP EXTENDETH HIT SELF IN ALTITVDE SPHERICALLYE AND THEREFORE
IMMOVABLE THE PALLACE OF FELICITYE GARNISHED WITH PERPETVALL SHININGE GLORIOUS LIGHTES INNVMERABLE
FARR EXCELLINGE OVR SONNE BOTH IN QVANTITYE AND QVALITYE THE VERY COVRT OF CELLESTIALL ANGELLES
DEVYD OF GREITE AND PERFECTE ENDLESSE IOYE THE HABITA-TLE FOR THE ELECT.
THE ORBE OF SATVRNE h MARINGE HIS REVOLVTION IN 30 YEARES.
THE ORBE OF IVPITER 24 MARINGE HIS PERIODE IN 12 YEARES
THE ORBE OF MARS ♂ MARINGE HIS REVOLVTION IN 2 YEARES
THE GREAT ORBE CARREINGE THE SPHERE OF THE SVNNE ☉ HIS CIRCLER PERIODE
THIS GLOBE OF MORALITYE DETERMINETH OVR YEARE.
THE ORBE OF VENVS ♀ REVOLVETH ROVNDE IN 9 MONETHS
THE ORBE OF MER ☿ CVRS IN 3 DAYES
THE SONNE

FIGURE 2.4. The Copernican planetary system. (From 16th Century drawing by Thomas Digges.)

but the observations were incorrect because of the limitations of the available instruments. It is now known that the apparent directions of the nearer fixed stars do in fact change against the background of more distant stars.

Tycho Brahe and Johannes Kepler

One of those who was interested in the work of Copernicus but could not accept his theory, largely because the directions of the stars seemed to remain unchanged, was the Danish astronomer Tycho Brahe. He died in 1601, before the telescope was invented, but during his lifetime Brahe developed several astronomical instruments. With these he was able to make more accurate observations than had been previously possible of the locations of the stars and planets in the sky. In particular, he made detailed records of the positions of the planet Mars during its nine apparitions from 1580 to 1600. These data, together with those obtained in the course of the next two opportunities, were destined to play an important role in the understanding of planetary motion.

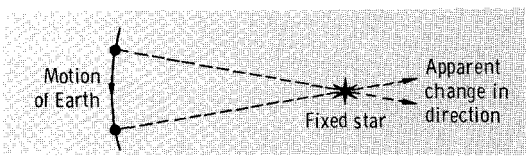


FIGURE 2.5. Change in direction of fixed star due to Earth's motion.

Two years before his death, Tycho Brahe

had moved to a location near Prague at the invitation of the German emperor, Rudolph II. There he was joined a year or so later, in 1600, by the young German astronomer Johannes Kepler who became Brahe's assistant and subsequently his successor as the imperial astronomer. In Kepler's first important work on cosmography, published in 1596, he stated his objections to the Ptolemaic system and gave reasons for favoring that of Copernicus in spite of its apparent defects. In this respect he was at variance with the man he was to serve a few years later. Tycho Brahe, in his Tychonic system, suggested that the five starlike planets revolved about the Sun, as in the Copernican theory, but he thought that the Sun and Moon moved around Earth, as in Ptolemy's system.

It is reported that, on his deathbed, Tycho Brahe asked Kepler to utilize the extensive observations that he, Brahe, had made on Mars to provide support for the Tychonic system. As an objective scientist, however, Kepler proceeded to use the data to test the three existing theories of planetary motion, namely those of Ptolemy, Copernicus, and Brahe. But none of these systems was able to account in a satisfactory manner for the recorded observations. After several years of tedious calculations on the motion of Mars, often made to six significant figures and without the benefit of logarithms, Kepler came to a revolutionary conclusion; this conclusion was to lead to the replacement of the earlier systems, with their circular deferents and epicycles, by one that was both much simpler and at the same time more accurate.

Kepler's Elliptical Planetary Orbits

For nearly 2000 years, the thinking of astronomers had been dominated by the Greek philosophical concept that the heavenly bodies could move only in perfect circles or in combinations of such circles. In the words of Aristotle, "the shape of the heavens is of necessity spherical . . . the circle is primary among figures, and the sphere occupies the same position among solids." Kepler, however, concluded from his study of Brahe's observations on Mars that the orbits of the planets are not circles but ellipses. By making this simple change, the complicated Ptolemaic system of deferents and epicycles was shown to be unnecessary and it was soon abandoned.

In his effort to understand the motion of Mars, Kepler started with two assumptions. His first was that, as in the Copernican system, Earth moves in a known circular orbit about the Sun. In such an orbit the distance from Earth to the Sun would always be the same. Furthermore, the time required for Earth to complete an orbit is exactly a year; that is, 365 days. Hence, if its rate of motion is uniform, as it should be in a circular orbit, the position of Earth in its orbit on any date can be determined.

The second assumption made by Kepler was that Mars also revolves about the Sun but in an unknown orbit. By noting the position of Mars relative to the Sun and to the fixed stars, it had been calculated that Mars requires 687 days to make a complete orbit around the Sun (p. 37). Consequently, for a certain location of Mars in the sky on any date, it was known that 687 days later it would return to exactly the same position. Because Earth has an orbital period of 365 days, it will be in different locations on these two dates, but the positions in its orbit will be known. From the quadrangle formed by the Sun, the two known positions of Earth in its orbit, and the particular location of Mars, as shown in figure 2.6, it is possible to calculate the distance of this planet from the Sun in terms of the constant distance of Earth from the Sun.

In this way, by utilizing Brahe's accumulated data and his own obtained in 1602 and

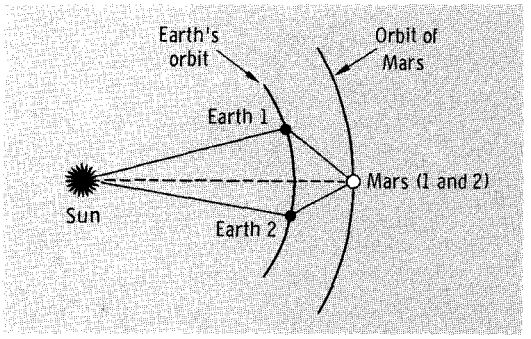


FIGURE 2.6. Determination of distance from Mars to the Sun.

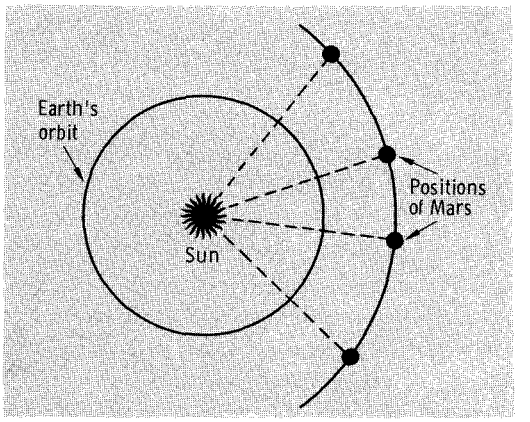


FIGURE 2.7. Determination of the orbit of Mars.

1604, Kepler was able to calculate the distance of Mars from the Sun for any specified position of the planet in the sky. Repetition of the same calculation for many different positions permitted Kepler to determine the distance of Mars from the Sun at each location and thus to define the orbit of the planet (fig. 2.7). Because the distances were found to vary with the position of Mars in its orbit, it was clear that the orbit could not be a circle with the Sun at its center. It might, however, have been a circle with the Sun displaced from the center. After much trial and error, Kepler was eventually forced to conclude that the orbit of Mars, and presumably of the other planets, is an ellipse.

Kepler's Laws of Planetary Motion

A circle is characterized by having a central point (or center) such that the distance from it to any point on the circle, that is, the radius, is a constant. An ellipse, on the other hand, has two points within it, called foci (plural of focus), so that the sum of the distances, known as the radius vectors, from them to any point on the ellipse is always the same. Thus, in figure 2.8, F_1 and F_2 are the two foci of the ellipse, and the sum of the distances PF_1 and PF_2 is the same no matter where the point P lies on the ellipse. According to Kepler's first law of planetary motion, published in 1609, the orbit of each planet is an ellipse with the Sun at one focus. Because the Sun is presumed to be stationary, all the planetary orbits must have one common focus, the location of the Sun.

Kepler's second law, enunciated at the same time as the first, was also based on a study of the orbit of Mars. This law, which has an important bearing on the length of the Martian seasons, as will be seen in chapter

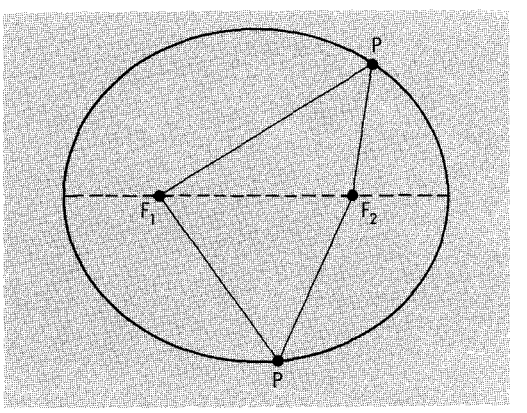


FIGURE 2.8. Properties of an ellipse.

III, states that a straight line joining the Sun and a planet in its orbit, that is, a radius vector of the ellipse, sweeps out equal areas in equal intervals of time. Thus, if the shaded portions of figure 2.9 have equal areas, it takes a planet the same time to cover the distances *AB* and *CD*. It is evident that when a planet is farther from the Sun, at *AB*, it covers a shorter distance in a given time, and hence travels more slowly, than when nearer to the Sun at *CD*. Kepler thought that the speed of a planet at any point in its orbit is inversely proportional to the length of the radius vector at that point. This is not true in general, however, but only at the two points when the planet is farthest and nearest, respectively, from the Sun.

The first two laws of Kepler did not, by themselves, account for the apparent occasional reversed or retrograde motion of the planets. But by taking into consideration the different orbital speeds of Earth and Mars, Kepler showed how the retrograde movement arises. A more detailed treatment of this subject is given in the next chapter. Another point that emerged was that the retrograde behavior should occur at regular intervals, once in each orbital period of a planet, as, in fact, it does.

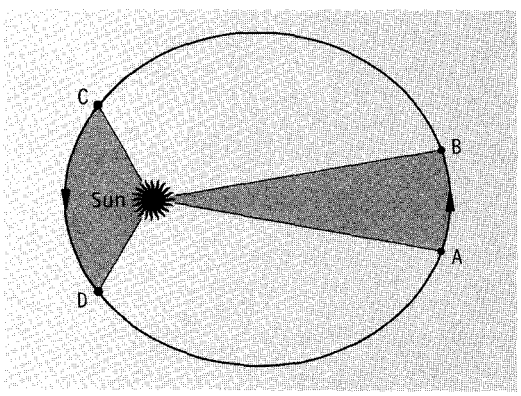


FIGURE 2.9. Illustration of Kepler's second law.

In 1618, Kepler completed his work with the third law of planetary motion; namely, that the square of the orbital period of a planet is proportional to the cube of its mean distance from the Sun. Although Kepler's three laws, based on a Sun-centered system, gave a much more accurate representation of planetary motion than did the Earth-centered system of Ptolemy, they still lacked a unifying principle. This came in 1687 when Isaac Newton, the English mathematician, published his theory of universal gravitation. By postulating that the Sun attracted each planet by a force that was inversely proportional to the square of its distance, the three laws of Kepler followed as a natural consequence. Thus the already dying Ptolemaic system was finally laid to rest, and the heliocentric system of Copernicus gained general acceptance. Earth was no longer the center of the Universe but merely one of the members of the Sun's system of planets.

It should be no detraction from Kepler's brilliant and painstaking work to mention that he was fortunate in several respects. The orbits of the planets around the Sun do not differ very greatly from circles. It happens that Earth's orbit is very close to being circular, as Kepler postulated, but the orbit of Mars deviates appreciably from a circle. If the reverse had been true, the determination of the Martian orbit might well have proved beyond Kepler's capability. Furthermore, the availability of the observational data on Mars, and Brahe's recommendation that Kepler make use of them to study planetary motion, was another fortunate circumstance. The orbit of the planet Venus is even closer to a circle than is that of Earth. Had this planet been selected for study instead of Mars, the discovery of Kepler's first and second laws would undoubtedly have been delayed.

Another interesting point has been noted by the Yale University astronomer Dirk

Brouwer. "An ellipse," he writes, "is only a first approximation to a planetary orbit. If an attempt were made to represent modern . . . observations with the aid of elliptic orbits, the inadequacy of the representation would soon be apparent. The good fortune of Kepler was that . . . [Brahe's] observations were just accurate enough to reveal the elliptical character of the [Martian] orbit. If they had been too accurate, the perturbations would have shown up, and Kepler's effort might well have bogged down in endless experimentations that would have led nowhere." The perturbations (or deviations) mentioned are attributable mainly to the gravitational attractions of Earth and Jupiter on Mars. If these (and other planetary) attractions were absent, the Martian orbit would be very close to an ellipse. For many purposes, however, such as those to be considered in this book, it is adequate to assume that Mars has an essentially elliptical orbit.

MARS THROUGH THE TELESCOPE

Early Telescopic Observations of Mars

The telescope, invented in 1608 and developed as an astronomical instrument by Galileo Galilei in Italy in the following year, provided support for the Copernican system. If the planets did indeed move in orbits about the Sun, then Venus, which is between Earth and the Sun, should show distinct crescent, gibbous, and full phases like the Moon. Because these phases could not be seen, at least by the unaided eye, it was argued that Venus does not travel around the Sun. When Galileo turned his telescope toward Venus, however, he was able to observe the different phases, eliminating the objection to the view that Venus orbits about the Sun.

The apparent shape of Mars was also expected to exhibit changes, but because this planet is farther from the Sun than Earth, such changes are relatively minor. Galileo looked for them and at the end of 1610 he wrote in a letter to a friend: "I dare not affirm that I was able to observe the phases of Mars; nevertheless, if I am not mistaken, I believe I have seen that it [Mars] is not perfectly round."

The earliest known drawing of Mars as seen in the telescope is shown in figure 2.10; it was made by Francisco Fontana in Italy in 1636. He said that "the disk of Mars is not uniform in color (*discolor*), but it appears fiery in the concave [sic] part." The nonuniform appearance of the surface may have been caused by details that Fontana was unable to resolve. The sketch is of historical interest only, and must not be taken seriously. The dark spot in the center, for example, was apparently caused by a fault in Fontana's telescope, as a similar spot appears in his drawing of the planet Venus. Furthermore, the ring near the edge of the disk has no counterpart in later telescopic observations of Mars.

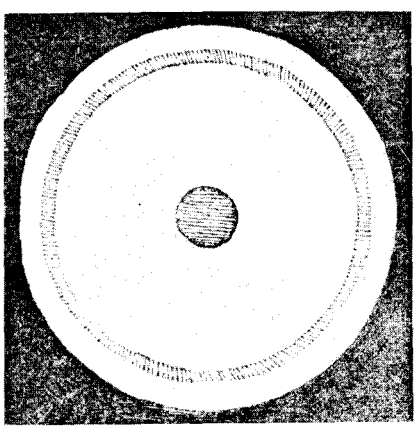


FIGURE 2.10. Earliest known drawing of Mars; made by Fontana in 1636. (From C. Flammarion, "La Planète Mars," Vol. 1.)

Fontana's drawings are of interest, however, because one made in August 1638, at the next apparition of Mars, shows a definite gibbous phase (fig. 2.11), as required by the Copernican (and Keplerian) theory. The dark central spot and the outer ring are, however, still present. Actually, the departure from the circular shape in the sketch is somewhat greater than it should be, but there seems little doubt that Fontana did observe and record the expected change in the shape of the visible part of the Martian disk.

In 1659, the Dutch mathematician and physicist Christian Huygens, who had improved the telescope by the use of better lenses, made the first recognizable drawing of Mars (fig. 2.12). It shows a conspicuous dark, triangular area, which was at one time called the Hourglass Sea (Mer du Sablier), since it looks like half of an hourglass. It is now known as Syrtis Major (fig. 2.21).

Before proceeding further with a review of the early studies of Mars, a short digression is necessary. The image of a celestial object as seen in a telescope is always inverted. It has been the universal practice among astronomers since the early 17th century to show sketches and photographs of such images in the way they are seen in the telescope; that is, in the inverted form. Such is the case in Huygens' drawing in figure 2.12, and also in nearly all the sketches and pictures in this book. The south of all illustrations of Mars is at the top and north at the bottom. The current practice is to refer to the right as west and the left as east, but this has apparently not always been so (fig. 2.13).

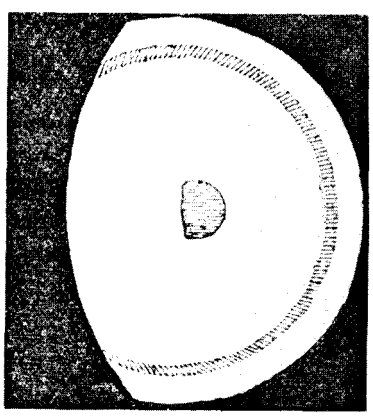


FIGURE 2.11. Fontana's drawing showing gibbous phase of Mars. (From C. Flammarion, "La Planète Mars," Vol 1.)

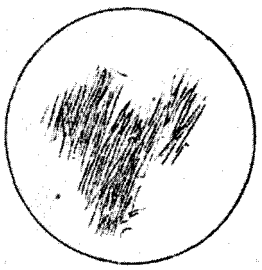


FIGURE 2.12. Earliest known drawing of Mars showing identifiable surface feature; made by Huygens in 1659. (From C. Flammarion, "La Planète Mars," Vol. 1.)

Because he could observe a particular distinctive feature on the Martian surface, Huygens was able to show that the planet rotates, like Earth, about a north-south axis. He found the period of rotation, that is, the length of a day on Mars, to be close to 24 terrestrial hours. Some 7 years later, in 1666, the Italian astronomer Giovanni D. Cassini, who is said to have discovered the rotation independently, made more precise observations and found the length of the Martian day to be 24 hours 40 minutes. This is about 2½ minutes longer than the value accepted today (p. 64).

Cassini is also supposed to have observed the polar caps of Mars, as indicated by the sketch in figure 2.13, where the letter *M* (Latin, *meridies*) stands for south and *S* (*septentrio*) for north. The drawing also shows white (or light-colored) areas on the

equator, which may represent some of the bright regions on the Martian surface. According to the Greek-born French astronomer, Eugenios M. Antoniadi, the upper and lower areas of the dark dumbbell-shaped feature which crosses the equator are the regions now known as Sinus Meridianii and Mare Acidalium; the line joining them is said to be Gehon, a so-called canal, but this seems questionable.

A more convincing indication of the south polar cap is seen in the somewhat crude sketch made by Huygens in 1672. This drawing (fig. 2.14) also shows the triangular Syrtis Major. A somewhat similar representation was made in 1719 by Cassini's nephew Giacomo F. Maraldi; he drew attention to the fact that the center of the cap did not coincide exactly with the apparent Martian pole. It may be noted that Earth's arctic and antarctic icecaps are also not centered on the poles. Maraldi reported that the appearance of the Martian surface changed while he was observing the planet in the telescope. W. Ley has suggested that this may have been caused by one of the occasional duststorms that occur on the planet (ch. VII).

Extensive reports with several sketches, based on observations of Mars made in 1777

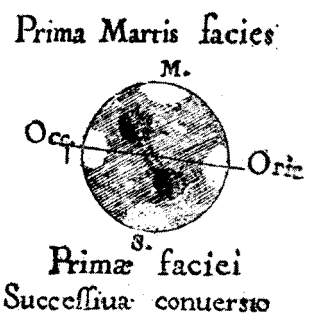


FIGURE 2.13. Drawing of Mars by Cassini showing polar caps. (From C. Flammarion, "La Planète Mars," Vol. 1.)

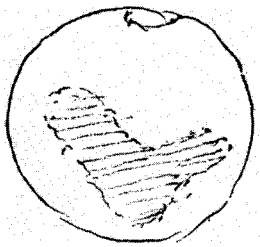


FIGURE 2.14. Drawing by Huygens showing south polar cap. (From C. Flammarion, "La Planète Mars," Vol. 1.)

and during subsequent opportunities, were published in 1781 and 1784 by William Herschel, the German-born musician turned telescope builder and astronomer, who was then private astronomer to George III of England. Among other things, he determined the angle between the axis of rotation of Mars and the perpendicular to the plane of its orbit around the Sun. Herschel found that this angle, about 30 degrees, was not very different from the corresponding angle of inclination of Earth's axis.

This angle of inclination determines the seasons on Earth (ch. III); hence, Herschel argued, Mars should also have four seasons. Furthermore, he thought that the planet "has a considerable but moderate atmosphere." Because the length of the day is also close to that on Earth, Herschel concluded that the "inhabitants [of Mars] probably enjoy a situation similar to our own." Incidentally, in the 17th and 18th centuries it was commonly accepted that Mars (and other planets) were inhabited.

Like several of his predecessors, Herschel considered the dark areas on the planet's surface to be oceans or seas, just as the dark areas on the Moon were regarded as being bodies of water. The lighter colored regions were believed to be dry land. It is now known, however, that such open bodies of liquid water cannot be present on Mars, and that

both dark and bright areas are essentially dry land. The color differences are probably caused by the nature of the surface materials.

The bright polar caps were assumed by Herschel to consist of snow and ice, like those of Earth. In this respect it is possible, but by no means certain, as will be pointed out in chapter VI, that he may have been right. Furthermore, he argued that the caps were probably not thick because they decrease considerably in size and almost disappear, as a result of melting and evaporation, during the Martian summer in each hemisphere.

One of Herschel's interesting observations, which he may have interpreted incorrectly, was that there were changes apparently occurring on the surface of Mars. "Besides the permanent spots," he wrote, "I have often noticed occasional changes in partial bright belts; and also once on a darkish one, in a pretty high latitude." Herschel attributed these changes to "clouds and vapors floating in the atmosphere of the planet." It is now known, however, that parts of the Martian surface undergo seasonal changes in appearance and these are generally most marked at high latitudes; that is, near the poles. There is a possibility, therefore, that the changes Herschel reported may not have been caused by clouds.

Maps of Mars

Although several astronomers had made drawings of certain aspects of Mars, there was no complete map of the planet until the early 1830's when one was compiled by the German banker Wilhelm Beer in conjunction with the astronomer Johann H. von Mädler. It was not a very good map and few of its features can be related to areas of Mars that are recognized today. An improved version, represented in figure 2.15, was produced in 1840. It shows the southern (left) and northern (right) hemispheres of Mars. The region indicated by the letter *f* is clearly Syrtis Major and *ae* is the Sabaeus Sinus region.

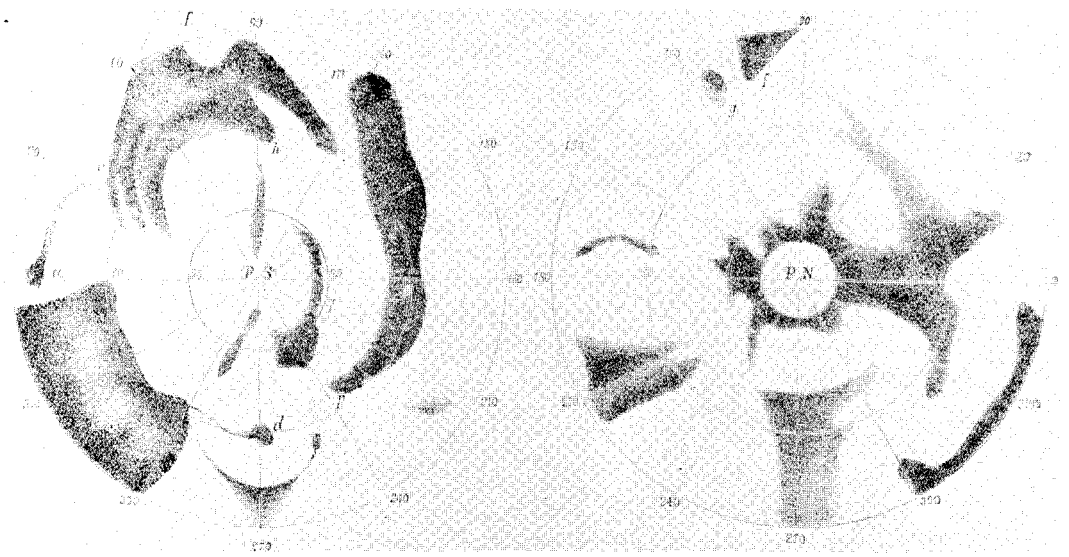


FIGURE 2.15. Map of Mars made by Beer and von Mädler in 1840. (From C. Flammarion, "La Planète Mars," Vol. 1.)

Some, but by no means all, of the other dark areas can also be related to those which appear on modern maps.

The French astronomer Camille Flammarion, in his classical work entitled "La Planète Mars et ses Conditions d'Habitabilité," published in 1892, called attention to a surprising error in the 1840 map of Beer and Mädler. The location of the northern tip *f* of Syrtis Major at the right of figure 2.15 is incorrect. It is shown between 90° and 100° (or so) of longitude, whereas it should have been between about 60° and 70°, as indicated in the map of the southern hemisphere.

An important aspect of the work of Beer and Mädler is that they were the first to define Martian (areographic) latitude and longitude. The zero of latitude must, of course, lie on the equator and this, by definition, is at all points equally distant from the north and south poles of the planet. Because the poles, through which passes the axis of rotation, have fixed positions, the equator (and zero latitude) is determined automatically.

The selection of the zero of longitude, that is, the prime (or zero) meridian, however, is quite arbitrary. On Earth, the prime meridian is chosen so as to pass through Greenwich, a borough of London, England, where the famous Royal Observatory is located. On Mars, Beer and Mädler arbitrarily selected the small dark area marked *a* in figure 2.15 as the point through which the prime areographic meridian should pass. The zero of longitude on Mars currently used is essentially the same as that introduced by Beer and Mädler. However, degrees of longitude are now numbered in the opposite direction around the planet. Thus, in a modern map, such as the one in figure 2.21, Syrtis Major is indicated between longitudes 280° and 295°. The corresponding longitudes in the system of Beer and Mädler would be obtained by subtracting these numbers from 360°, yielding 80° and 65°.

After the middle of the 19th century, the sketches of Mars produced by various astronomers improved markedly, possibly because of the availability of better telescopes. Some good representations were drawn in 1858 by the papal astronomer Father Pierre Angelo Secchi in Rome. In 1863 he published the first known sketches of the planet in color, and these led him to suspect that the color of the Martian surface changed from time to time. This suspicion was confirmed 12 years later by Camille Flammarion.

Some excellent drawings of Mars were made in 1864 by the Reverend William R. Dawes in England. One of these is reproduced in figure 2.16, partly because of its interest in another connection, as will be explained shortly. The dark region of the figure shows clearly and with considerable accuracy Sabaeus Sinus, at the left, then the twin points of Meridianii Sinus, followed by Margaritifer Sinus and Aurorae Sinus in current terminology. The area at the bottom is presumably Mare Acidalium. Incidentally, Dawes

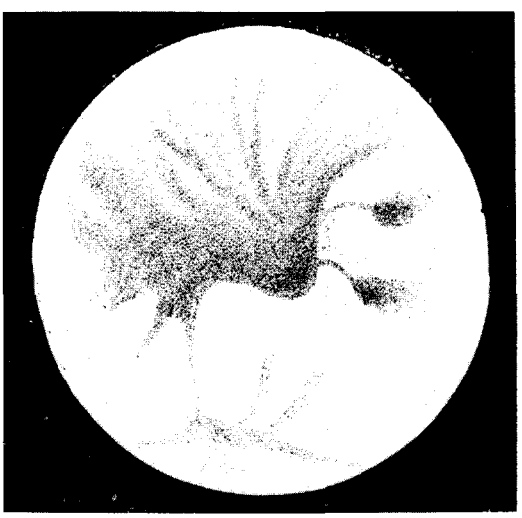


FIGURE 2.16. Drawing of Mars made by Dawes in 1864. (From C. Flammarion, "'La Planète Mars," Vol. 1.)

thought that, because of the clarity and constancy of the surface markings and the rarity of clouds, Mars had a "thin" atmosphere.

Naming the Features of Mars

The sketches of Mars made by several observers, including those of Secchi and Dawes, mentioned above, of N. Lockyer (1862), F. Kaiser (1862, 1864), and of others, were combined in 1867 by the well-known English astronomer, science writer, and lecturer Richard A. Proctor in the map shown in figure 2.17. In this map, Proctor made the first systematic attempt to name the characteristic features of the Martian surface as an aid to their identification. Most were named after various observers of the planet; for example, Tycho Brahe (Tycho Sea), Kepler (Kepler Land), Beer (Beer Sea), Mädler (Mädler Land), and Dawes (Dawes Ocean).

It may be noted that the zero of longitude, where the east and west hemispheres of Proctor's map touch each other, corresponds as closely as possible to the prime meridian of Beer and Mädler. The degrees of longitude are also numbered in the same manner; thus, Syrtis Major, called Kaiser Sea by Proctor, is seen between longitudes of approximately 70° and 90°.

An improved map of Mars, using Proctor's nomenclature, was published in England in 1877 by Nathaniel E. Green, but in general the proposed names were not received with any degree of enthusiasm. A possible reason may be found in the objections expressed by Flammarion. "Fault has been found," he wrote, "with Proctor for having given too much recognition to the astronomers of his own country, and for having repeated the same names." He pointed out, for example, that the name of Dawes was applied to no less than six features—ocean, continent, sea, strait, isle, and bay—and that this was a source of confusion. On the other hand, there were many eminent astronomers and other scientists whose names

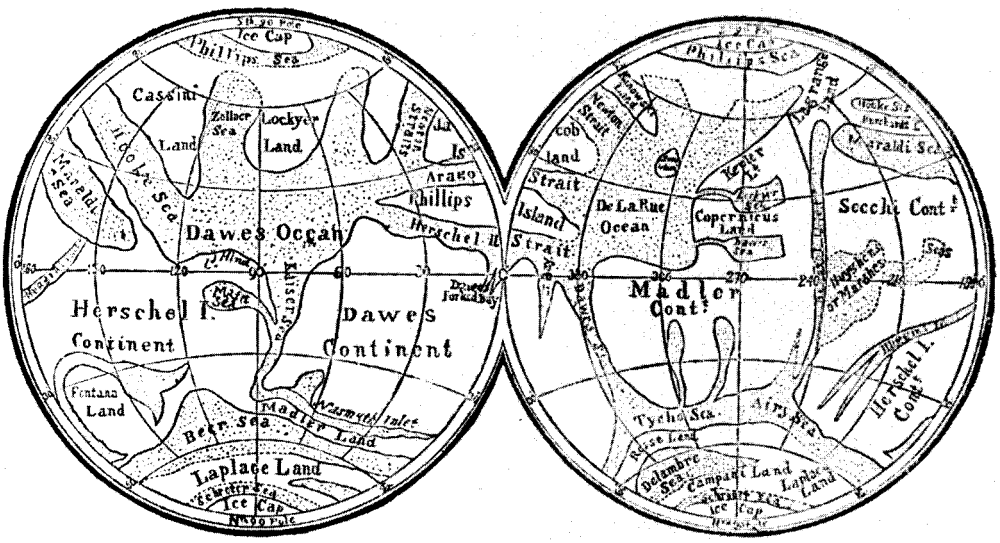


FIGURE 2.17. Proctor's map of Mars with named features. (From C. Flammarion, "La Planète Mars," Vol. 1.)

might well have been used. In 1876, Flammarion drew a map of Mars and introduced his own nomenclature, in French, based on that of Proctor but avoiding duplications. This, however, was also destined to become of only historical interest.

For reasons that will be given in chapter III, there are certain times, at intervals of 15 or 17 years, when Earth and Mars are exceptionally close together. These occasions are of special interest to astronomers because Mars then exhibits its maximum apparent size and brightness. The conditions are consequently particularly favorable for observing the planet. Such a close approach occurred in 1877 and that year was marked by two events of historical significance in the study of Mars.

The first, which will be considered more fully in chapter IV, was the discovery by the American astronomer Asaph Hall that Mars has two moons or satellites. The report of this discovery was received without argument, and the moons of Mars soon became an accepted part of astronomy. The other event, described below, related to the so-called canals; it was to initiate a heated controversy that cannot be regarded as completely resolved more than 90 years later.

One observer, who took advantage of a moderately good telescope and the clear Italian skies during the favorable apparition of Mars in the summer of 1877, was Giovanni Virginio Schiaparelli, director of the Brera Observatory, Milan, Italy. In the first report of his studies of Mars, he recorded, with much greater accuracy than had previous observers, the latitude and longitude of 62 characteristic features on the surface of the planet. Furthermore, he disregarded Proctor's nomenclature, as well as that of Flammarion of which he may not have been aware, and introduced a completely new one.

The names used by Schiaparelli were in Latin and were drawn from geography and mythology. The bright areas of the planet, called "lands" and "continents" by Proctor, were named after terrestrial countries, either real, such as Arabia, Hellas (Greece), Syria, etc., or mythical, such as Elysium, Amazonis, Utopia, etc. The dark areas, also following Proctor, were designated seas; e.g., Boreum Mare (North Sea), Tyrrhenum Mare (Tyrrehenian Sea), etc. Schiaparelli also identified a number of bays, as did Proctor, such as Sabaeus Sinus (Sabian Bay) and Aurorae Sinus (Aurora Bay); some large bays he called gulfs, as for example, Golfo Sabeo. Several small dark areas were designated lakes; thus, Solis Lacus (Lake of the Sun) and Niliacus Lacus (Egyptian Lake). The system of Latin nomenclature introduced by Schiaparelli in 1877 received general approval and he and others added more names in later years. With some modifications, the system is still in common use (fig. 2.21).

In ascribing coordinates of latitude and longitude to the various regions on Mars, Schiaparelli used a zero point essentially identical with the ones employed by his predecessors, Beer and Mädler and Proctor. This point, where the prime meridian intersects the equator, he called Fastigium Aryn, after the name of the mythical village shown on medieval maps as located on the cupola of Earth. In Schiaparelli's maps of Mars, the prime meridian passed through the center of Meridianii Sinus (Meridian Bay), and this was accepted until quite recently.

The "Canals" of Mars

Although Schiaparelli was responsible for some of the best contemporary maps of Mars and was the first to locate many of the features of the planet with considerable accuracy, he is probably best remembered in connection with his observations of canals. On Schiapa-

relli's maps there appear a number of dark, almost straight lines, several of them being many hundreds of miles in length. These are seen clearly in figure 2.18, which is a map of Mars prepared from observations made at five apparitions of Mars between 1877 and 1886. Schiaparelli emphasized, however, that only a relatively small number of the lines can be seen at any one instant (p. 25).

The remarkable linear features were named after rivers; for example, Indus, Ganges, Hiddekel (Tigris), etc. In his textual description, however, Schiaparelli referred to them as *canali* (plural of *canale*), the Italian word for channels. There is little doubt that he originally considered the *canali* to be wide, natural waterways connecting the various dark regions which he thought were bodies of water. The word "canale" was, however, translated as "canal" in most European languages, and this led to an erroneous impression concerning what Schiaparelli had intended the word to mean. In order to be visible from Earth, under the best conditions of telescopic viewing, the canals would have to be several miles wide. Even if they contained water, which they do not, they would hardly be called canals in the sense that the word is commonly used.

It is frequently stated that Schiaparelli not only discovered the canals, but was also responsible for introducing the term *canale*. But neither of these statements is strictly correct. The word *canale* was used about 1869 by Father Secchi, to whom reference has already been made. There is evidence that Secchi and Schiaparelli exchanged information about Mars, and Schiaparelli must have thought of *canale* as an accepted term because he introduced it in a casual manner in his report of the observations he made in 1877.

As far as the present writer has been able to determine, the first mention of *canale* by Schiaparelli is in his report entitled "Astronomical and Physical Observations on the Axis of Rotation and on the Topography of the Planet Mars," published in Rome in 1878. On page 45 of his collected "Works," in the description of the second "horn" (or arm) of the Golfo Sabeo, he says: "This is formed by the inner tip [or extremity] of a gulf ending in a sharp point and is well defined. The *canale* which originates there at an opening is not easy to see, and some uncertainty may arise concerning it." Later, on page 167, he wrote: "Where the . . . peninsulas are connected to the two continental zones, they are flanked by the mouths of broad *canali* which cross these zones."

Moreover, Schiaparelli was not the first to identify linear features on Mars, as he himself pointed out. In a footnote to a report to the Society of Italian Spectroscopists, published in Palermo in 1878, but which may have been added later, Schiaparelli states: "Some people have been inclined to doubt the existence of the *canali* since they have not seen them. Yet many of these *canali* are not new and have already been seen by such excellent observers as Kaiser, Lockyer, Secchi, Green, etc." Schiaparelli might also have added the names of Dawes and Proctor, as the linear features are evident in the excellent drawings of the former, such as the one in figure 2.16, and in the rather crude map of the latter (fig. 2.17). Furthermore, the sketches made by the German astronomer Johannes Schroeter between 1785 and 1802, but which were not published until 1881, also showed similar lines.

Then, why is the discovery of the Martian canals so frequently attributed to Schiaparelli? The reason is twofold. First, he identified the *canali* by giving them specific names (of rivers) and, in so doing, called attention to their existence as a special aspect of the planet. Second, he recorded a much larger number of the linear features than had been

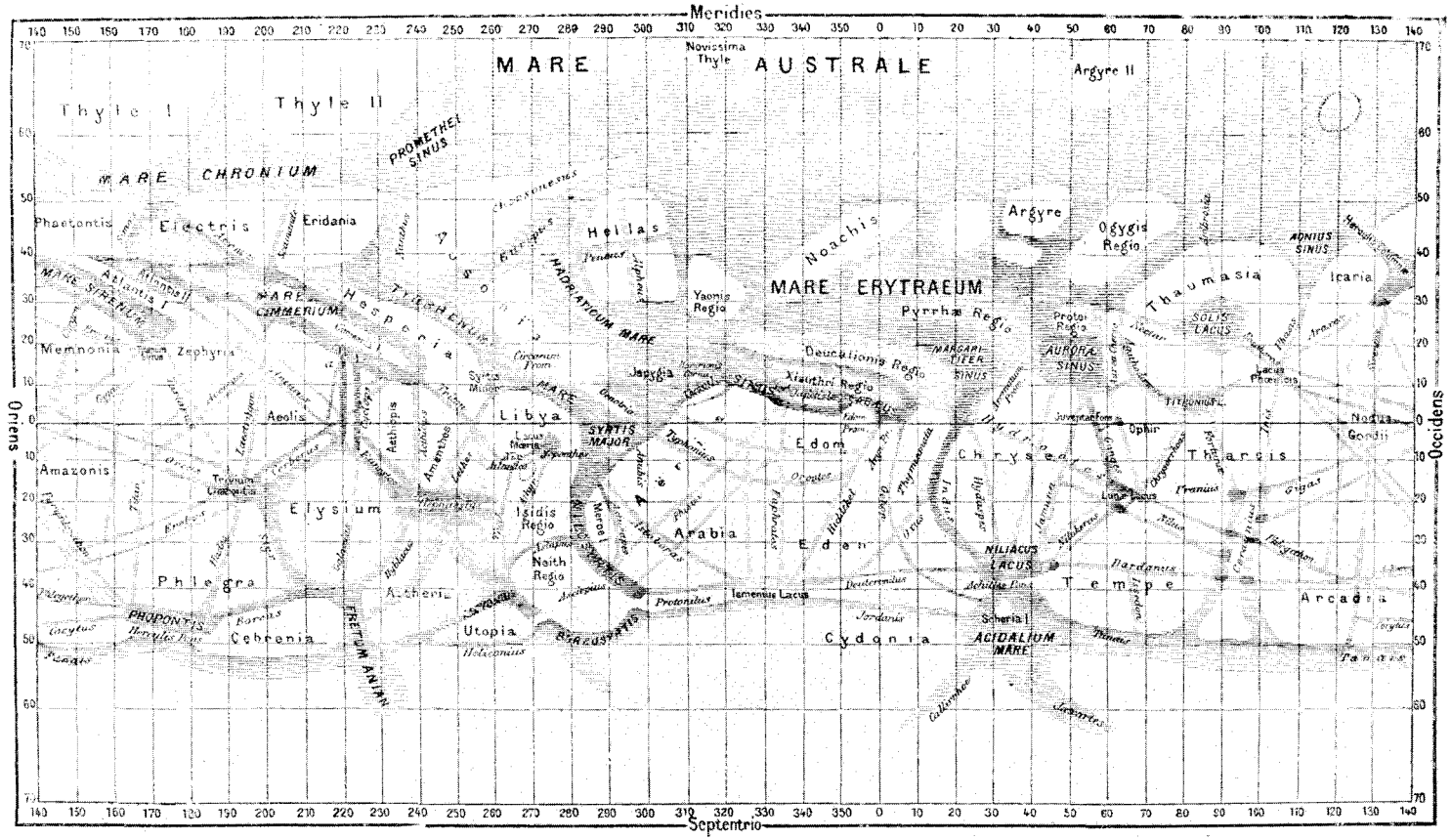


FIGURE 2.18. Mercator projection map of Mars by Schiaparelli based on observations made at five oppositions in 1877 through 1966. (From C. Flammarion, "La Planète Mars," Vol. 1.)

seen by any (or all) observers in the preceding 200 years. In the report to the Society of Spectroscopists mentioned above, Schiaparelli wrote, in italics for emphasis: "There are no large continuous continental masses on Mars, as the entire surface of the planet is divided by many *canali* into an enormous number of islands." This is immediately apparent from the map in figure 2.18, where the white areas are assumed to be land, and the dark portions were evidently regarded by Schiaparelli as being seas, lakes, or rivers (*canali*).

During the apparition of Mars in 1879, Schiaparelli claimed he saw canals that had not been apparent to him at the preceding opportunity in 1877. Furthermore, he reported the first instance of gemination or twinning in which a canal appeared as two parallel lines instead of a single line. The sketch in figure 2.19, made by Schiaparelli in 1888, shows several examples of the gemination of canals.

No other astronomers were able to see the numerous Martian canals until the 1886 opportunity when observers in the United States, England, and France claimed to have seen them. The delay in confirmation may possibly have been caused by the fact that the northern hemisphere of Mars, where Schiaparelli saw most of the canals (fig. 2.18), was more clearly visible from Earth during 1886 than it had been in the intervening apparitions of the planet.

Lowell and the Martian Canals

The problem of the canals on Mars would have been left to the scientists had it not been for the activities of Percival Lowell, a mem-

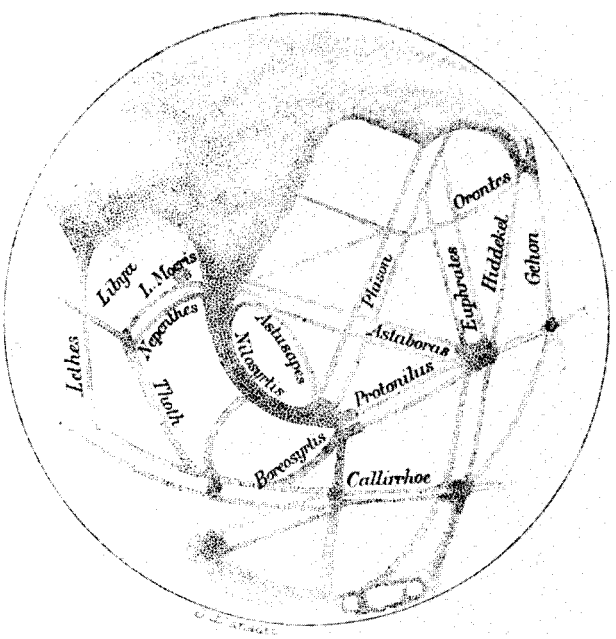


FIGURE 2.19. Drawing by Schiaparelli showing gemination of canals. (From C. Flammarion, "La Planète Mars," Vol. 1.)

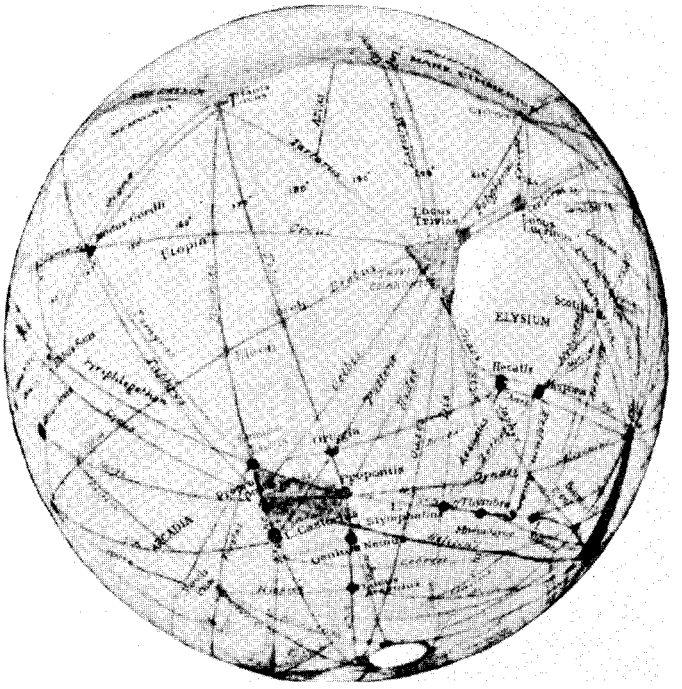


FIGURE 2.20. Drawing of Mars by Lowell showing canal network. (From P. Lowell, "Mars as the Abode of Life," Macmillan.)

ber of a distinguished and wealthy family in Boston, and a writer on the Far East. Lowell had become interested in planetary astronomy, especially in Mars, and to pursue this interest, he founded the Lowell Observatory near Flagstaff, Ariz., in 1894, to take advantage of the high altitude and clear atmosphere. Within a short time, Lowell had observed the Martian canals and claimed to have confirmed the phenomenon of gemination. As the years went by he reported more and more canals, and his maps of Mars were essentially networks of canals. This is evident in one of Lowell's sketches reproduced in figure 2.20.

The black spots shown in the drawing, almost invariably at the junctions of canals, were first reported in 1892 by William H. Pickering at the Harvard Observatory at Arequipa, Peru. He called them lakes, but Lowell, who identified almost 200 in all, referred to them as "oases," which he considered to be a more accurate description of their nature.

Altogether, Lowell identified over 500 canals. But out of fairness it should be mentioned that neither he nor Schiaparelli, who reported 113 canals, claimed that all the canals could be seen at any given time. For example, Schiaparelli said that the map in figure 2.18 "does not at all correspond to the appearance of Mars at any given period, because generally only a few [canals] are visible at one time." In his more flowery style, Lowell wrote: "Permanent the canals are in place, impermanent they prove in character. . . . Some will show when others remain hid and others will appear when the first have become

Features of Mars and Their Locations

Name	Long. (deg.)	Lat. (deg.)	Name	Long. (deg.)	Lat. (deg.)
Acidalium Mare	30	+45	Claritas	110	−35
Aeolis	215	−5	Copaïs Palus	280	+55
Aeria	310	+10	Coprates	65	−15
Aetheria	230	+40	Cyclopia	230	−5
Aethiopis	230	+10	Cydonia	0	+40
Amazonis	140	0	Deltoton Sinus	305	−4
Amenthes	250	+5	Deucalionis Regio	340	−15
Aonius Sinus	105	−45	Deuteronilus	0	+35
Arabia	330	+20	Diacria	180	+50
Araxes	115	−25	Dioscuria	320	+50
Arcadia	100	+45	Edom	345	0
Argyre	25	−45	Electris	190	−45
Arnon	335	+48	Elysium	210	+25
Aurorae Sinus	50	−15	Eridania	220	−15
Ausonia	250	−40	Erythraeum Mare	40	−25
Australe Mare	40	−60	Eunostos	220	+22
Baltia	50	+60	Euphrates	335	+20
Boreum Mare	90	+50	Gehon	0	+15
Boreosyrtis	290	+55	Hadriacum Mare	270	−40
Candor	75	+3	Hellas	290	−40
Casius	260	+40	Hellespontica Depressio	340	−60
Cebrenia	210	+50	Hellespontus	325	−50
Cecropia	320	+60	Hesperia	240	−20
Ceraunius	95	+20	Hiddekel	345	+15
Cerberus	205	+15	Hyperboreus Lacus	60	+75
Chalce	0	−50	Iapigia	295	−20
Chersonesus	260	−50	Icaria	130	−40
Chronium Mare	210	−58	Isidis Regio	275	+20
Chryse	30	+10	Ismenius Lacus	330	+40
Chrysokeras	110	−50	Jamuna	40	+10
Cimmerium Mare	220	−20	Juventae Fons	63	−5

Features of Mars and Their Locations (cont.)

Name	Long. (deg.)	Lat. (deg.)	Name	Long. (deg.)	Lat. (deg.)
Laestrygon	200	0	Propontis	185	+45
Lemuria	200	+70	Protei Regio	50	−23
Libya	270	0	Protonilus	315	+42
Lunae Palus	65	+15	Pyrrhae Regio	38	−15
Margaritifer Sinus	25	−10	Sabaeus Sinus	340	−8
Memnonia	150	−20	Scandia	150	+60
Meroe	285	+35	Serpentis Mare	320	−30
Meridianii Sinus	0	−5	Sinaï	70	−20
Moab	350	+20	Sirenum Mare	155	−30
Moeris Lacus	270	+8	Sithonius Lacus	240	+45
Nectar	72	−28	Solis Lacus	90	−28
Neith Regio	270	+35	Styx	200	+30
Nepenthes	260	+20	Syria	100	−20
Nereidum Fretum	55	−45	Syrtis Major	290	+10
Niliacus Lacus	30	+30	Tanaïs	70	+50
Nilokeras	55	+30	Tempe	70	+40
Nilosyrtis	290	+42	Thaumasia	85	−35
Nix Olympica	130	+20	Thoth	255	+30
Noachis	330	−45	Thyle I	180	−70
Ogygis Regio	65	−45	Thyle II	230	−70
Olympia	200	+80	Thymiamata	10	+10
Ophir	65	−10	Tithonius Lacus	85	−5
Ortygia	0	+60	Tractus Albus	80	+30
Oxia Palus	18	+8	Trinacria	268	−25
Oxus	10	+20	Trivium Charontis	198	+20
Panchaïa	200	+60	Tyrrhenum Mare	255	−20
Pandorae Fretum	340	−25	Uchronia	260	+70
Phaethontis	155	−50	Umbra	290	+50
Phison	320	+20	Utopia	250	+50
Phlegra	190	+30	Vulcani Pelagus	15	−35
Phoenicis Lacus	110	−12	Xanthe	50	+10
Phrixi Regio	70	−40	Yaonis Regio	320	−40
Promethei Sinus	280	−65	Zephryia	195	0

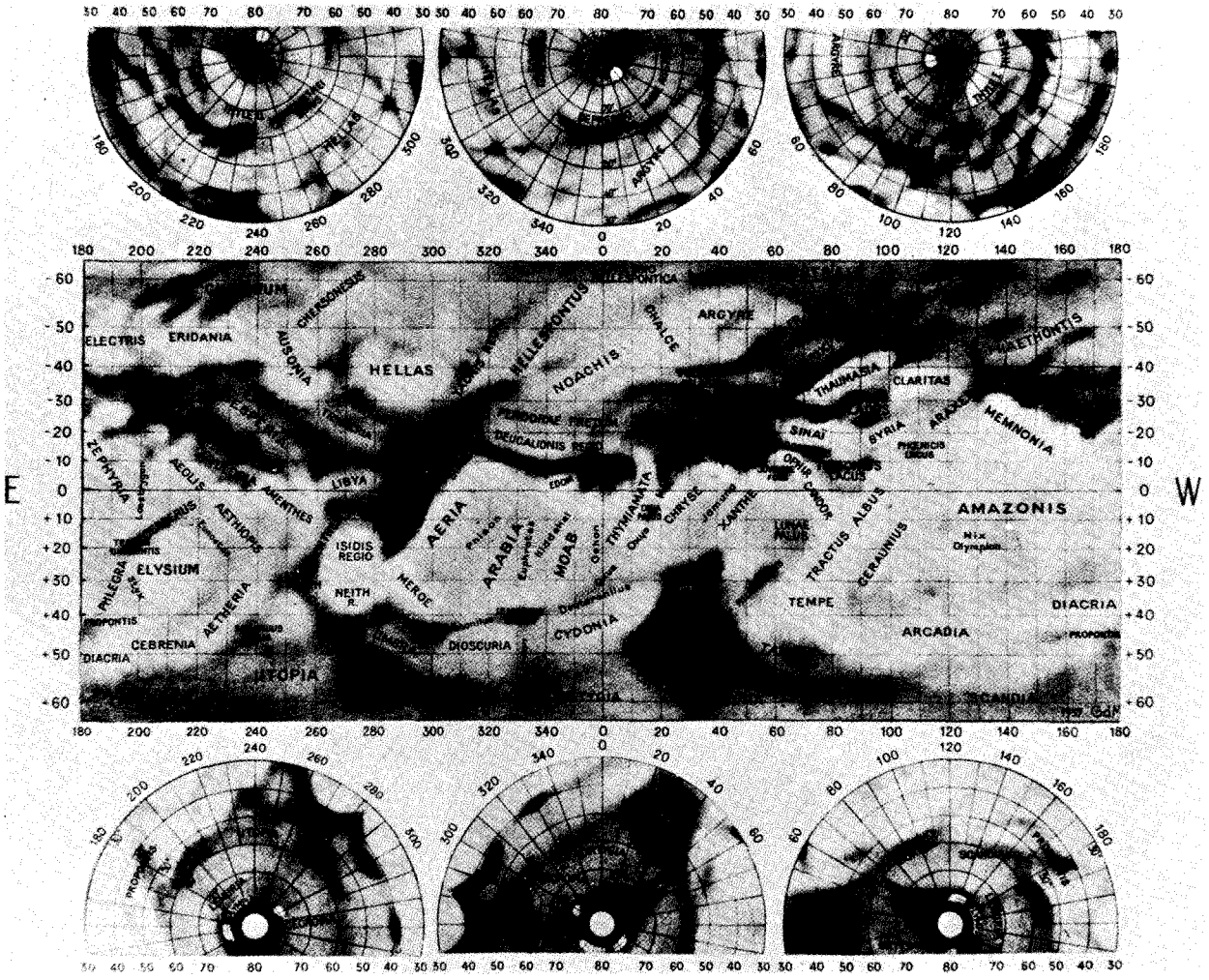


FIGURE 2.21. The International Astronomical Union map of Mars based on numerous telescopic observations.

invisible." Lowell also observed that the canals exhibit seasonal darkening in appearance, which he attributed to the development of vegetation as water became available from the polar caps.

The main reason for the interest taken by the general public in Lowell's observations on Mars was his firm opinion that the canals were not natural features but the work of "intelligent creatures, alike to us in spirit, though not in form." In three books, entitled "Mars," published in 1895; "Mars and Its Canals," in 1906; and "Mars as the Abode of Life," 1908, he developed this idea with increasing enthusiasm. Lowell realized that, as seen in the telescope, the canals had considerable width, too wide to be actual waterways. The dark lines, he contended, represented agricultural regions irrigated by water flowing from the Martian polar caps along true canals that are not actually visible. The oases, which like the canals exhibit seasonal variations in appearance, were regarded by Lowell as centers of population surrounded by areas of vegetation.

It is of interest that Schiaparelli, the "father" of the Martian canals, adopted a somewhat equivocal attitude concerning life on the planet. In an article published in 1893, before Lowell started his observations, Schiaparelli said of the canals: "It is not necessary to suppose them to be the work of intelligent beings, and . . . we are now inclined to believe them to be produced by the evolution of the planet." Later in the same article, when referring in particular to the double canals, he wrote: "Their singular aspect . . ., as if they were the work of rule and compass, has led some to see in them the work of . . . inhabitants of the planet. I am very careful not to dispute this supposition which includes nothing impossible." However, he went on to say, "The intervention of intelligent beings might explain the geometrical appearance of the gemination, but is not at all necessary for such a purpose." Then, in 1897, perhaps under the influence of Lowell, he wrote: "[The] arrangement [of the canals] presents an indescribable simplicity and symmetry which cannot be the work of chance."

Lowell's views concerning life on Mars created much excitement at the turn of the century. At the present time no scientist believes that there are intelligent beings on the planet. There is no general agreement, however, concerning the canals. Many reputable observers have claimed that they saw them, whereas others have stated equally firmly that, in spite (or because) of excellent seeing conditions, they have been unable to detect the canals. Currently, most astronomers would probably agree that there are linear features on the surface of Mars, but whether they are such long, thin straight lines, as Schiaparelli, Lowell, and others have reported seeing or not is still a matter of controversy. The problem is discussed more fully in chapter VI, but no final solution is offered.

Identification of Martian Features

To the 62 surface features on Mars which Schiaparelli identified and named in 1877, he later added more, as did other astronomers in subsequent years. Although the main aspects of the map of Mars remained unchanged, by the middle of the 20th century nearly 600 features had been named, in accordance with the scheme introduced by Schiaparelli. Unfortunately some newly discovered markings were given different names by different observers. Furthermore, certain named features were obviously temporary; they may have been seen on one occasion but not in later years.

Consequently, the International Astronomical Union appointed a committee to review the naming of the important and permanent aspects of the Martian surface and to prepare a map of the planet. A list was pub-

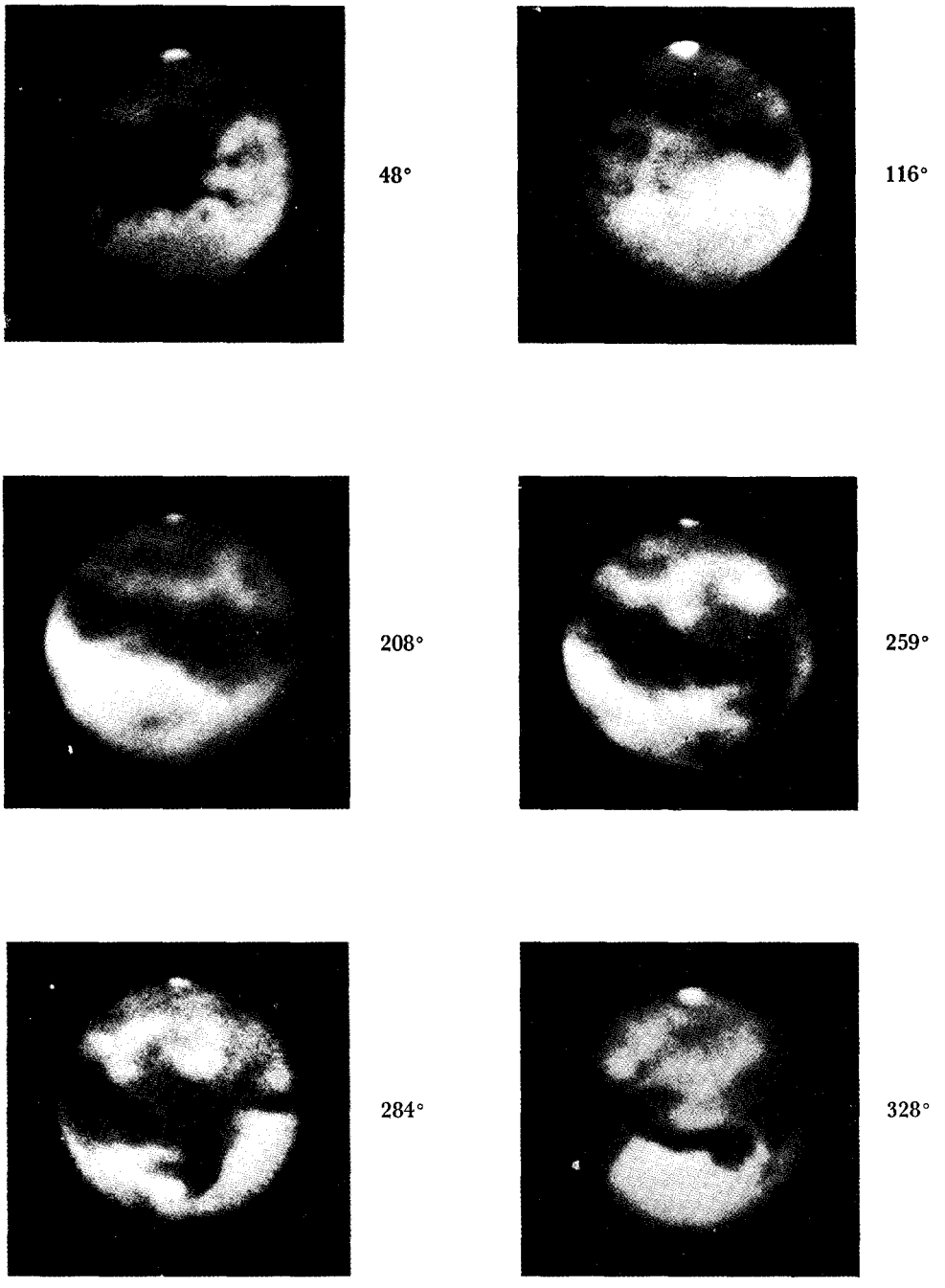


FIGURE 2.22. Complete circuit of Mars; the longitude of the central meridian is indicated in each case. (Lowell Observatory photographs.)

lished in 1958 of the names and locations, in latitude and longitude, of 128 apparently permanent features. These were given earlier in this chapter and the related map, drawn by G. de Mottoni, is presented in figure 2.21. The latter is a Mercator-type projection which exaggerates the longitudinal dimensions in the higher latitudes. Hence, the map covers the region between latitudes 65° N and 65° S only. The areas of high latitude are shown in the six polar views at the top (south) and bottom (north) of the chart.

The six photographs in figure 2.22, taken at the Lowell Observatory during the favorable opportunity presented in 1941, show a complete circuit of Mars. The latitude of the central meridian is indicated in each case. It is of interest to compare these photographs with the hand-drawn map in figure 2.21.

Recent observations on Mars, especially of its rotation, show that the prime (or zero) meridian, which must pass through both north and south poles, does not coincide with the one postulated by Schiaparelli and that had been in use for some 60 years. The middle of the Meridianii Sinus, to which Schiaparelli assigned 0° longitude, is now at a longitude of 357°. The new zero of latitude and longitude happens to be the northern tip of the western point of Meridianii Sinus.

FOR FURTHER INFORMATION

ANTONIADI, E. M.: La Planète Mars. Libraire Scientifique Herman et Cie, 1930.

FLAMMARION, C.: La Planète Mars et ses Conditions d'Habitabilité. Gauthier-Villars et Fils, vol. I, 1892; vol. II, 1909.

LEY, W.; VON BRAUN, W.; AND BONESTELL, C.: The Exploration of Mars. The Viking Press, 1956. (The bibliography in this book contains many references of historical interest.)

LOWELL, P.: Mars and Its Canals. Macmillan, 1906.

PICKERING, W. H.: Mars. R. C. Badger, 1921.

VAUCOULEURS, G. DE: Physics of the Planet Mars. Faber & Faber, 1954. (Contains many references of historical interest.)

319-889 O—69——2

III

Mars in the Solar System

THE SOLAR SYSTEM

Planets and Asteroids

THE PLANET MARS IS A MEMBER of the solar system; that is to say, it is one of the bodies which revolve in orbits about the Sun. During the early years of the 17th century, when Kepler's views of a Sun-centered system were beginning to find acceptance, six planets were known. In order of increasing distance from the Sun, they were Mercury, Venus, Earth, Mars, Jupiter, and Saturn.

These bodies are all visible, at appropriate times, by the unaided eye. In spite of the development of the telescope, they remained the only known planets until 1781. In that year, however, William Herschel identified the planet Uranus, beyond Saturn. Herschel's discovery was made by careful observation of the skies, but the two remaining planets Neptune and Pluto were discovered, in 1846 and 1930, respectively, as a result of mathematical calculations of the perturbations of the orbit of Uranus.

In addition to the nine planets, the solar system contains a group of small bodies, known as asteroids or planetoids, which also orbit around the Sun. More than 30 000 of these bodies, with diameters ranging from 760 kilometers (470 miles) down to a few kilometers (1 or 2 miles), have been observed in the telescope. The orbits of some 1600 of the asteroids have been determined, and the majority are found to lie between the orbits of Mars and Jupiter. This region has therefore become known as the asteroid belt.

A comparison of the orbits of the planets shows that there is an exceptionally large gap between Mars and Jupiter (fig. 3.1). In fact, Kepler had suggested that a previously unknown planet might be found in this region of the solar system. This conclusion appeared to find support from a rule based on the work of the German astronomer J. D. Titius, but brought to public notice by his fellow countryman J. E. Bode in 1771. By this rule, the distances of the successive planets from the Sun, in terms of the distance of Earth from the Sun, could be represented by adding 0.4 to each of the following series of numbers: 0, 0.3, 0.6, 1.2, 2.4, 4.8, and 9.6. Apart from the zero, these numbers

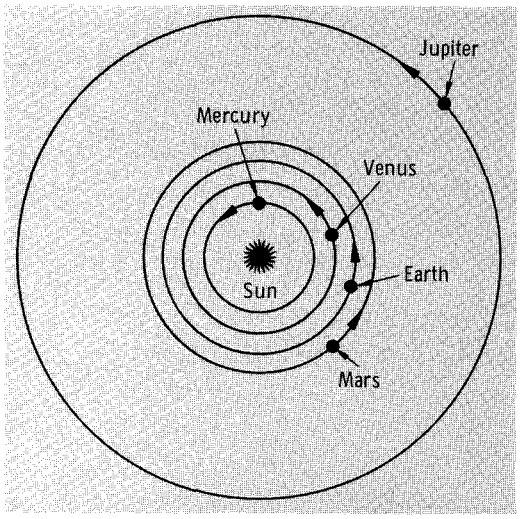


FIGURE 3.1. The solar system showing planetary orbits.

form a simple geometric progression, in which the ratio of successive terms is two.

Bode's rule, sometimes inappropriately called Bode's law, was found to correlate the distances of the six planets known before 1781, except that there was no planet with a distance corresponding to the number 2.4. Such a distance would lie between the orbits of Mars and Jupiter. When Uranus was discovered and its orbit determined, its distance from the Sun was found to agree with Bode's rule, using the next number in the geometrical series; namely, 19.2. Consequently, it appeared that there should be a planet between Mars and Jupiter, and at the end of the 18th century a group of astronomers made plans to look for this missing planet.

Before a systematic search began, however, the first asteroid was discovered by an Italian, Giuseppe Piazzi, in 1801. He observed a small object moving against the background of fixed stars and thought it might be a comet. But the famous German mathematician Karl F. Gauss calculated its orbit and showed that it was similar to the orbits of the planets. The average distance from the Sun of this new member of the solar system, which was called Ceres, appeared to be in good agreement with that required for the missing member, 2.4, of the Bode series.

For a time, it was thought that Ceres, with its orbit between those of Mars and Jupiter, was the missing planet. But by 1807, three other similar bodies had been found with orbits in the same general region. No further asteroids were observed until 1845 when a fifth was added, and subsequently the number has increased steadily almost from year to year. The smallest observable asteroids have diameters of a few kilometers, but there is little doubt that there are many more of smaller size which have not been detected. These very small members of the solar system, which are relatively easily deflected from their orbits by collisions, have probably played a role in shaping the surface of Mars (p. 131).

The asteroids may have resulted from the disruption or explosion, perhaps in the early stages of development of the solar system, of a planet between Mars and Jupiter. It is possible, on the other hand, that the considerable mass of Jupiter, the largest and most massive of the planets, may have prevented the formation of a single planet by accumulation of material in the nebular cloud from which the solar system evolved. Many small planetoids would then have formed instead.

Elliptical Orbits

Kepler's contention that the orbits of the planets (and asteroids) are elliptical has been confirmed. Although there are important minor deviations (or perturbations), they can be ignored here. It will be recalled from chapter II that the characteristic of an ellipse is that it has two focal points (or foci); the sum of the distances from any point on

the ellipse to the two foci is the same for all points on the ellipse. The line XY passing through the foci F_1 and F_2 in figure 3.2 is called the major axis of the ellipse. It can be easily shown that the constant sum of the distances to the two foci mentioned above is equal to the length of the major axis.

Consider, for example, the point X; this is a point on the ellipse and the distances to the foci are XF_2 and XF_1. The foci are located symmetrically within the ellipse; consequently, the distance XF_1 is equal to F_2Y. Hence the sum of the distances from X (or Y) to the two foci is XF_2+F_2Y, which is equal to the length XY, the major axis of the ellipse. By definition, the sum of the distances from any point on the ellipse to the two foci is always the same; hence, this sum is equal to the length of the major axis.

Another important property of an ellipse is known as the eccentricity. It provides an indication of the departure of the shape from that of a circle. If O in figure 3.2 is the midpoint of the major axis of the ellipse, then XO, which is half the length of the major axis, is called the semimajor axis. The eccentricity of the ellipse is defined by

$$\text{eccentricity}=\frac{OF_1}{OX} \quad \left(\text{or } \frac{OF_2}{OY}\right).$$

If an ellipse is imagined to become less and less elliptical, and approach a circle, the two foci must come closer together. In a circle, the foci coincide at the center of the circle; then the points F_1, O, and F_2 in figure 3.2 are all identical. The distance OF_1 is zero and hence, by the definition given above, the eccentricity of a circle is zero. On the other hand, if an ellipse becomes more elliptical, and deviates more and more from a circle, the foci F_1 and F_2 get farther apart. The distance OF_1 then increases (relative to OX) and so does the eccentricity.

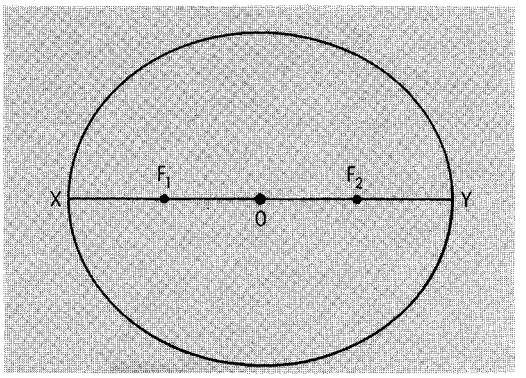


FIGURE 3.2. Characteristics of an ellipse.

In the solar system, the Sun is stationary at one focus of a number of ellipses; the other focus, which is different for each ellipse, is unoccupied. The orbits of the planets, as well as of the asteroids, are each represented by one of these ellipses. Each orbital ellipse has its own particular major (or semimajor) axis and eccentricity which completely define its shape and size. The length of the major axis of the orbit of Mars is 455.88 million kilometers (283.27 million miles) and the eccentricity is 0.0933. The latter may be compared with 0.0167 for the eccentricity of Earth's orbit and 0.00679 for the orbit of Venus. The eccentricity of the Martian orbit is, in fact, greater than that of any of the planets other than Mercury, the closest to the Sun, and Pluto, the farthest. As stated in chapter II, the small eccentricity of Earth's orbit, which does not differ greatly from a circle, and the relatively large eccentricty of the orbit of Mars were important factors in enabling Kepler to discover the first two laws of planetary motion.

An exaggerated elliptical orbit around the Sun is indicated in figure 3.3. Planetary orbits have very much smaller eccentricities, but the sketch might represent the orbits of some asteroids, several of which have fairly large eccentricities (up to about 0.3). The points

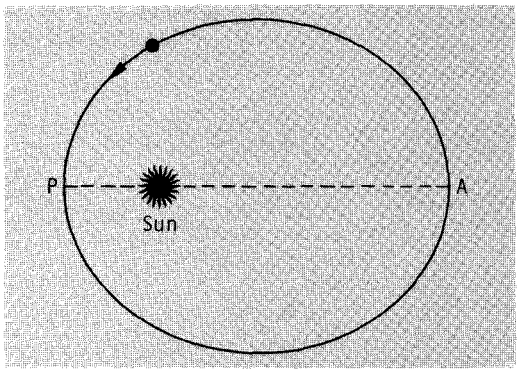


FIGURE 3.3. Aphelion and perihelion of a planetary orbit.

to be considered are independent of the eccentricity, however, and a highly elliptical shape is used for clarity.

In the figure, the major axis is now represented by the line *PA*. At the point *P*, the planet (or asteroid) is the closest it ever gets to the Sun; this point is called the perihelion, from *peri* meaning nearby, and *helios* for Sun in Greek. The distance from the Sun to *P* is called the perihelion distance. Similarly, the point *A* where the planet is farthest from the Sun, is known as the aphelion, from *apo* meaning away from, and the distance from *A* to the Sun is the aphelion distance. The more highly eccentric the ellipse, the greater is the difference between (or rather the ratio of) the perihelion and aphelion distances.

For Mars, the distance of closest approach to the Sun, the perihelion distance, is 206.66 million kilometers (128.41 million miles), whereas the farthest (or aphelion) distance is 249.22 million kilometers (154.86 million miles). The difference is thus quite significant; namely, about 42.5 million kilometers (26.45 million miles). For comparison purposes, the aphelion and perihelion distances of Earth are 152.2 million kilometers (94.56 million miles) and 147.0 million kilometers (91.34 million miles). These numbers show that the orbit of Earth does not depart greatly from a circle, but that of Mars does.

In accordance with Kepler's second law (p. 14), a planet moving in an elliptical orbit does not travel at constant speed. Its speed is greatest at perihelion, when it is closest to the Sun, and least at aphelion, when it is farthest away. At these two points, the speeds are inversely proportional to the respective distances from the Sun. For Mars, the maximum orbital speed, at perihelion, is 26.4 km/sec (16.4 mps), and the minimum, at aphelion, is 22.0 km/sec (13.6 mps). The average speed of the planet over the whole of its orbit is 24.13 km/sec (14.99 mps). Earth's average orbital speed is 29.8 km/sec (18.5 mps). The greater speed with which Earth travels in a smaller orbit than that of Mars has interesting consequences, as will be seen shortly.

Sidereal and Synodic Periods

The time required by a planet to make a complete orbit (or complete revolution) about the Sun is the sidereal year for that planet. The adjective sidereal, derived from the Latin, *sidus,* meaning a celestial body or constellation, is used to distinguish the sidereal year from the synodic year, which is the apparent time of revolution of a planet as seen from Earth. The difference between the two kinds of year may be explained in the following manner, which also accounts for the use of the term synodic, derived from the Latin *synodus* (or Greek *synodos*) for meeting.

Figure 3.4 shows the orbit of a planet around the Sun and also a distant star (not a planet) *S*. The latter is so far away from the Sun that its position may be regarded as fixed. The point *X* is that of the planet at some position in its orbit when it is directly in line with the Sun and the particular fixed

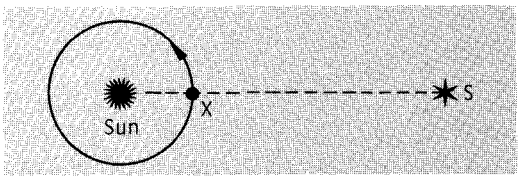


FIGURE 3.4. Definition of the sidereal year.

star S. The time elapsing between two successive occasions of this kind is the sidereal year; that is, the actual time it takes the planet to make one complete orbit around the Sun. For Mars, the sidereal year is 686.980 (essentially 687) Earth days, compared with 365.256 days for Earth's sidereal year. The greater length of the Martian year is caused by the longer orbit and its lower average speed, as given above.

The synodic year, better called the synodic period, is the time between two successive similar alinements (or meetings) of the planet with the Sun as seen from Earth. Let figure 3.5 represent the orbits of Earth and of another planet such as Mars, which is farther out from the Sun. Suppose that Mars is in the position M relative to Earth at E; the Sun, Earth, and Mars are then in a straight line (or, more correctly, in the same plane), with Earth between the Sun and Mars. In this location, Mars is said to be in opposition, because, as seen from Earth, Mars is opposite the Sun. On the other hand, if Mars is at M' relative to Earth at E, when the Sun is directly between Earth and Mars, the planet is in conjunction.

The synodic year is then defined as the time elapsing between two successive meetings of the same kind, that is, either two oppositions or two conjunctions, of Mars, Earth, and the Sun. Because Mars cannot be observed at conjunctions, as it is then in the same direction as the Sun, the length of the synodic year is defined, from the practical standpoint, as the time between two successive oppositions.

Although Mars, Earth, and the Sun are alined (or, at least, in one plane) at each opposition, the actual positions of the planets in their orbits change from one opposition to the next. A typical situation is depicted in figure 3.6 which shows the locations of Earth and Mars at two successive oppositions, indicated by the numbers 1 and 2. Earth is at E_1 (and Mars at M_1) at the first opposition and at E_2 (Mars at M_2) at the next opposition. Between the two oppositions, Earth has made two complete orbits around the Sun, and, in

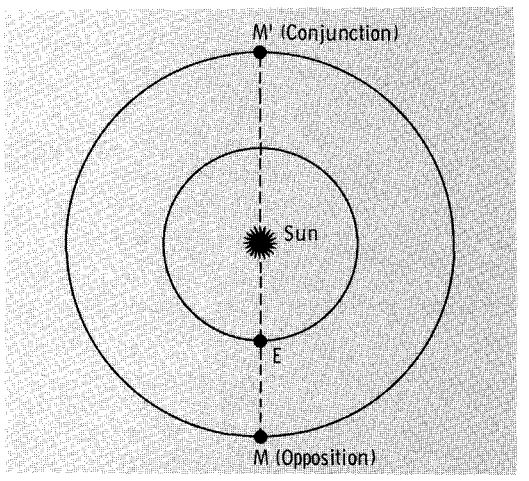


FIGURE 3.5. Mars at conjunction and opposition.

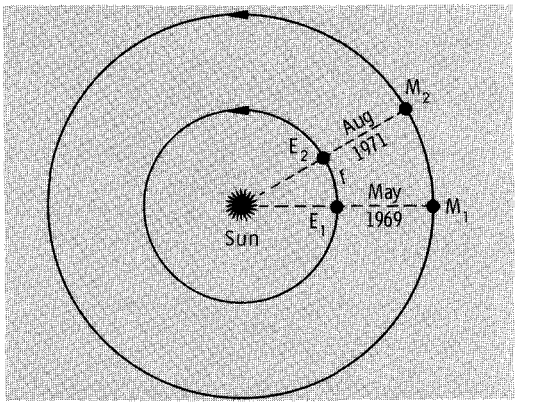


FIGURE 3.6. Locations of Earth and Mars at successive oppositions.

addition, has traversed the distance from E_1 to E_2. Mars, on the other hand, with its considerably longer orbit and lower speed, has made only one complete orbit plus the distance from M_1 to M_2. The locations of E_2 and M_2, relative to E_1 and M_1, respectively, are determined by the lengths of the orbits and the orbital speeds of the two planets.

Suppose the distance from E_1 to E_2 represents a fraction f of Earth's orbit. Then the distance from M_1 to M_2 will be the same fraction f of the orbit of Mars. This would be exactly true only if the two orbits had the same shape. Although it is known that they do not, it will be assumed for the moment that the distances E_1E_2 and M_1M_2 are both the same fraction f of the respective orbits. Consequently, between one opposition and the next, that is, during the same time interval, Earth travels a distance of $2+f$ Earth orbits, whereas Mars covers a distance of $1+f$ Mars orbits. The time required for Earth to make a complete orbit of the Sun is 365 days (ignoring fractions), whereas Mars requires its sidereal year of 687 days to complete one orbit. The synodic period of Mars, which is the time between two oppositions, can thus be expressed in two alternative ways; in terms of Earth's motion

$$\text{Synodic period of Mars} = (2+f) \times 365 \text{ days}$$

whereas in terms of the motion of Mars,

$$\text{Synodic period of Mars} = (1+f) \times 687 \text{ days}$$

In order to find the value of f, these two expressions are set equal to each other; thus

$$(2+f) \times 365 = (1+f) \times 687$$
$$f = 0.134$$

Consequently, the synodic period of Mars is given either by 2.134×365, or 1.134×687. In each case the result is 780 days.

According to the foregoing calculation, oppositions of Mars should occur every 780 days; that is, at intervals of about 2 years and 50 days. Because the orbits of Earth and Mars do not have the same shape, as indicated by the marked difference in the eccentricities, the fraction f is not the same for both orbits. In fact, the values vary with the positions in the orbits where the oppositions occur. Consequently, the observed intervals between successive oppositions actually range from 763 to 810 days. The average value, however, has been determined to be 779.935 days, in very good agreement with the 780 days calculated above.

OBSERVATIONS OF MARS

Favorable and Unfavorable Oppositions

For making approximate estimates, the time between oppositions may be taken as roughly 2 years and 7 weeks. It can now be seen why the apparitions of Mars occur at intervals of a little over 2 years (p. 5). It is at these times that Mars and Earth are closest and the conditions are most favorable for observations of Mars. A contributory factor is that the planet is then directly opposite from the Sun and this provides the best conditions of visibility.

Because the eccentricity of the orbit of Mars is significantly greater than that of Earth's orbit, the distance between the two planets varies from one opposition to the next. This is apparent from figure 3.7 which gives the locations of Earth and Mars in their respective orbits, at several oppositions, and also shows the major axis of the orbit of Mars. The distance between the two planets ranges from somewhat less than 56 million kilometers (35 million miles) at the most favorable opposition to a little over 101 million kilometers (63 million miles) at the least favorable. The favorable oppositions, when Mars and Earth

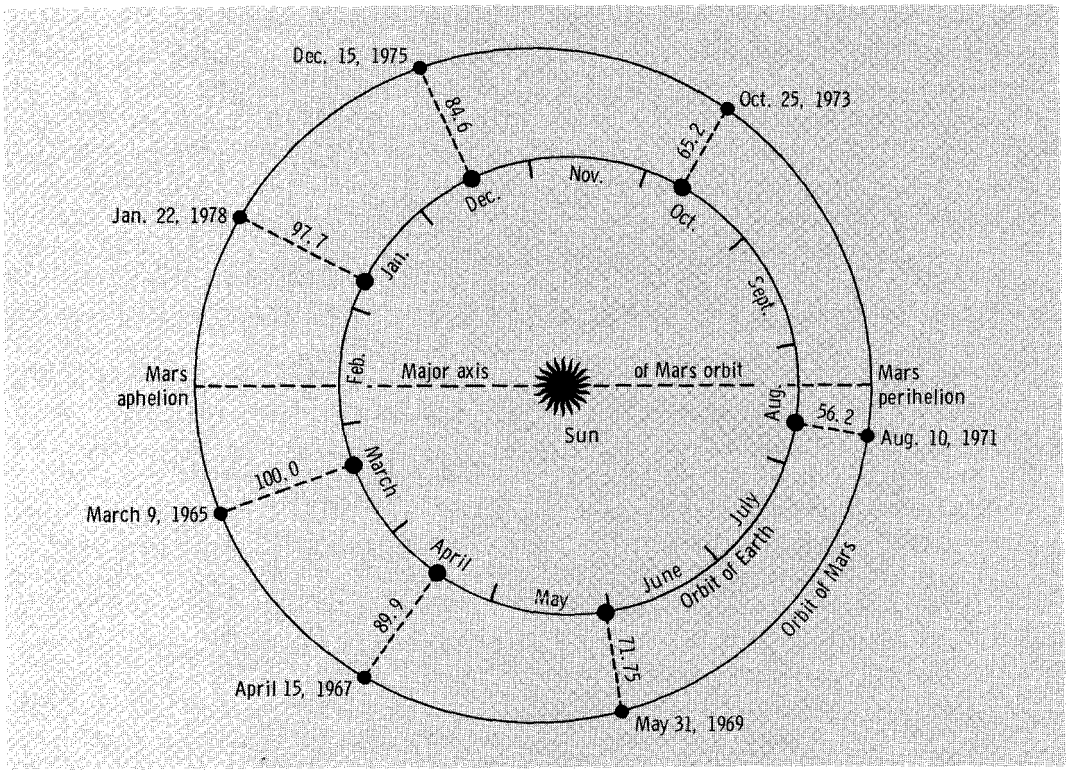


FIGURE 3.7. Earth and Mars at several oppositions; distances are in millions of kilometers.

are closest, occur when Mars is in the region of its perihelion, whereas the unfavorable ones are around aphelion. This is a consequence of the marked eccentricity of the Martian orbit. The eccentricity of Earth's orbit is too small to be significant on the scale of figure 3.7 and so it appears to be a circle with the Sun at the center.

The dates (in months) at which Earth is at various locations of its orbit are included in the figure. It is seen that the most favorable (perihelic) oppositions always take place around July, August, and early September and the least favorable (aphelic) ones are generally in January, February, and early March. Although the perihelion oppositions are best for observing Mars, intermediate oppositions, in the months of May, June, October, and November, are also quite good because the distances to Earth are never more than about 76 million kilometers (47 million miles).

For purposes of reference, the dates of the oppositions from 1969 through 1980 are given in the accompanying table, together with the respective distances between Mars and Earth. It will be observed that the opposition of August 10, 1971, is a particularly favorable one. Although it is often tacitly assumed that Earth and Mars approach most closely at the time of opposition, this is not quite true. Actually, the closest approach may occur from a few days before to a few days after the date on which Mars and the Sun are exactly oppo-

site as seen from Earth. The dates of closest approach for each opposition are included in the table given below.

If the orbits of Earth and Mars were exactly parallel, the time of closest approach would coincide with opposition. But the orbits are not parallel, partly because of the different eccentricities and partly because they do not lie in the same plane. The inclination of the orbital planes of the two planets will be considered shortly.

The most favorable oppositions, when Earth and Mars are within about 56 to 58 million kilometers (35 to 36 million miles) of each other, occur at intervals of 15 or 17 years. It is seen in figure 3.6 that between two successive oppositions, the planets advance by a fraction f in their orbits. The value of f is 0.134, on the average, and this is between one-seventh (0.143) and one-eighth (0.125). It follows, therefore, that it would require between seven and eight oppositions for the planets to return to a particular location in their orbits. Because oppositions must be counted in whole numbers, this means that during every seventh or eighth opposition the planets will be approximately in the same region of their orbits.

The average time between successive oppositions is somewhat less than 26 months; hence, seven oppositions represent roughly 15 years and eight oppositions about 17 years. Consequently, the most favorable (perihelic) oppositions, which will, however, not be identical, will occur at intervals of either about 15 or about 17 years. The last favorable opposition, for example, occurred in 1956 (September 10), and the next will be in 1971 (August 10). The interval is close to 15 years in this case. There were similar intervals between the successive favorable oppositions of 1909, 1924, and 1939. On the other hand, the interval was 17 years between the aphelic oppositions of 1892 and 1909, and of 1939 and 1956.

An exact repetition of the dates (and orbital locations) of oppositions could occur only when a complete number of Martian (sidereal) years was exactly equal to a complete number of Earth years. After such an interval, the two planets would return to precisely the same positions in their respective orbits. A very close approximation to this situation arises every 284 Earth years (103 733 days), which is almost exactly equivalent to 151 Martian years (103 734 days). Hence, the terrestrial dates of oppositions, within a day or so, recur after a lapse of 284 years.

Mars Oppositions and Closest Approaches

Opposition date	Closest approach	Distances	
		Million kilometers	Million miles
May 31, 1969	June 9, 1969	71.75	44.58
Aug. 10, 1971	Aug. 12, 1971	56.20	34.92
Oct. 25, 1973	Oct. 17, 1973	65.22	40.53
Dec. 15, 1975	Dec. 8, 1975	84.60	52.57
Jan. 22, 1978	Jan. 19, 1978	97.72	60.72
Feb. 25, 1980	Feb. 26, 1980	101.32	62.96

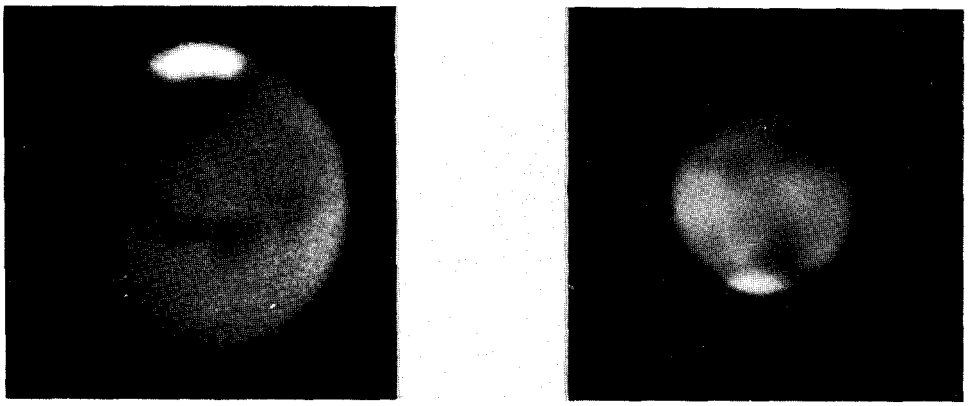


FIGURE 3.8. Photographs of Mars taken at favorable (left) and unfavorable (right) oppositions. (Lowell Observatory photographs.)

The apparent diameter of Mars, as seen by the eye or, better, through a telescope, is inversely related to its distance from Earth. The ratio of the distances between the two planets at the most favorable and least favorable oppositions is roughly 35/63. The ratio of the apparent diameters of Mars is 63/35, or 1.8, so at the most favorable oppositions Mars appears to be 1.8 times as large as at the least favorable oppositions, as seen from Earth. The photographs of Mars in figure 3.8 were taken in 1924 (left) and 1931 (right) at favorable and unfavorable oppositions, respectively.

Another factor which makes it easier to see Mars at the favorable oppositions is that the planet is then close to perihelion; that is to say, it is at its closest to the Sun. The brightness of Mars, like that of the Moon, depends on reflected sunlight. Consequently, when Mars is closest to the Sun, at perihelion, it receives and reflects more light than at aphelion. In other words, in addition to appearing larger, Mars is also brighter at favorable than unfavorable oppositions.

Because Earth has a smaller orbit than that of Mars and also travels faster, Earth catches up with Mars before opposition and moves away from it after opposition. As a result, there is a period of only a few months each side of an opposition when the conditions are best for studying Mars. Apart from distance, there are other considerations that restrict observations of the planet to the times around opposition. Because oppositions occur, on the average, at intervals of more than 2 years and favorable ones at intervals of 15 and 17 years, it is evident that the opportunities for making detailed studies of Mars are quite limited.

Eastern and Western Quadratures

In addition to the phenomena of opposition and conjunction, there are two other configurations of Mars with reference to Earth and the Sun that are of interest. In figure 3.9 certain positions of Mars in its orbit during a synodic period are shown relative to the position of Earth. Actually, the diagram is a composite of several in which both Mars and Earth have changed their locations, but it may be supposed that these are superimposed in such a way that Earth is always at the same point. Thus, the figure shows some succes-

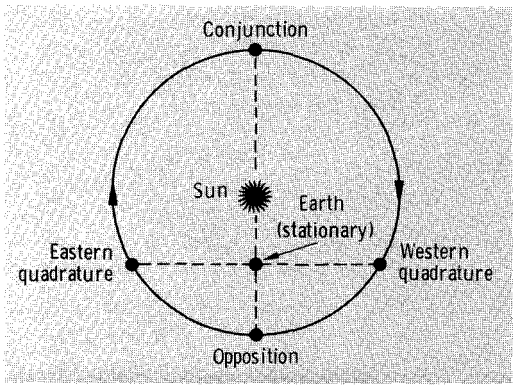


FIGURE 3.9. Eastern and western quadratures.

sive positions of Mars, in relation to the Sun, as seen from Earth during a complete synodic period, that is, from one opposition (or conjunction) to the next.

As viewed from the north ecliptic pole (fig. 3.17), all planets travel in their orbits in a counterclockwise direction around the Sun (fig. 3.1). Because the orbital speed of Mars is less than that of Earth and its orbit is larger, it appears from Earth as if Mars is moving in the opposite (or clockwise) direction, as shown in figure 3.9. The situation is similar to that in which, to an observer on a fast-moving train, a train traveling more slowly on an adjacent track appears to be going backward.

The term quadrature refers to a quarter of a circle, that is 90 degrees of arc. Thus, at eastern quadrature the planet Mars appears from Earth to be 90 degrees east of the Sun. In other words, the direction of Mars is at an angle of 90 degrees east of the direction of the Sun. Similarly, at western quadrature, the direction of Mars is 90 degrees west of the Sun. The angle between the directions of the Sun and a planet as seen from Earth is called the elongation. Hence, at the quadratures the elongation is 90 degrees (east or west); at conjunction the elongation is zero and at opposition it is 180 degrees.

Daily Rising and Setting of Mars

From figure 3.9 some ideas can be obtained concerning the times when Mars is visible and where it can be seen in the sky. First something must be said of the apparent motion of the Sun and the planets (and in fact of all celestial bodies). The Sun may be regarded as being stationary and the planets do not travel very far in the course of a single day. Their apparent daily (or nightly) motion from east to west is thus merely a consequence of Earth's rotation from west to east. Because the orbits of the planets lie in planes not very different from Earth's orbital plane, the planets follow the same general path across the sky as does the Sun. Thus, the planets rise in the east, like the Sun, and set in the west. The times of rising and setting are, however, generally quite different from those of the Sun.

At (or near) conjunction, Mars is in the direction of the Sun and it cannot then be seen because of the Sun's brilliance. Soon after conjunction, Mars appears to be just west of the Sun (fig. 3.10*A*), and it should be observed rising low in the eastern sky just before dawn. After sunrise, although Mars is still west of the Sun, the planet disappears from view because of the brightness of the sky. Mars is then said to be a morning star. It is so far from Earth, however, and is visible for only a short time near the horizon, that the conditions for studying the planet are very poor.

On subsequent days, Mars rises farther and farther west of the Sun; that is to say, the planet can be seen in the east above the horizon for longer and longer periods before sunrise. At western quadrature (fig. 3.10*B*), Mars would be 90 degrees from the Sun; con-

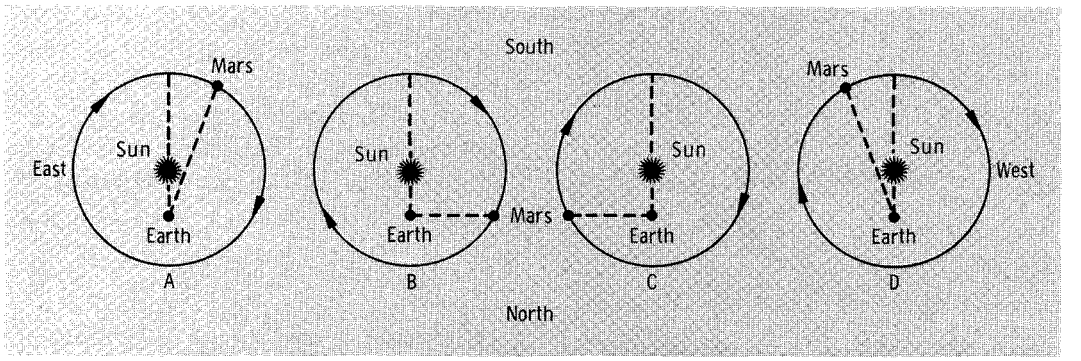


FIGURE 3.10. Positions of Mars, relative to Earth and the Sun, at various times.

sequently, in the early morning, when the Sun is toward the east, Mars would be toward the south. The planet would then be visible during the latter part of the night until the sky becomes bright enough to interfere. When opposition is reached, Mars is at 180 degrees from the Sun. Because the Sun is in the south at midday, Mars will then be in the south at midnight. The planet can be observed for several hours each night, between sunset and sunrise. Provided Mars rises sufficiently above the horizon (p. 48), the conditions are then best for viewing the planet.

After opposition, Mars appears in the south earlier and earlier (and sets earlier in the west) each night. At eastern quadrature the planet is 90° east of the Sun (fig. 3.10*C*); hence, at sunset, when the Sun is in the western sky, Mars will be in the south. It can then be observed during the early part of the evening because it will have set by midnight. Following eastern quadrature, Mars moves westward relative to the Sun (fig. 3.10*D*) and is then called an evening star because it is visible only for a few hours after sunset. In due course, the planet works its way around to conjunction again having made a complete cycle in the synodic period of 780 days.

It was mentioned in chapter II that Mars exhibits a partial (or gibbous) phase. The proportion of the surface of the planet that is visible from Earth depends on the relative positions of the Sun, Earth, and Mars. At conjunction and opposition, the full face of Mars is illuminated by the Sun; at these times the planet's disk appears (or would appear if it could be seen) as a complete circle. At all other times, part of Mars is in shadow, just as the Moon is when it is not full.

The reason is that, although one hemisphere of the planet is always presented to the Sun, whereas the other hemisphere is in shadow, the illuminated half cannot be seen in its entirety when the directions of Sun and Earth, with reference to Mars, are not the same. The situation is illustrated in figure 3.11; only the portion *AB* of the illuminated part of the surface of Mars is visible from Earth, whereas the part *BC*, which should be seen, is in shadow.

The angle made at Mars by the directions of the Sun and Earth is known as the phase angle of the planet, as shown in the sketch. It can be seen that this angle is equal to the angle between *B* and *C*. Because the whole disk of the planet, represented by *AC*, makes an angle of 180 degrees, it is evident that the fraction of the disk that is in shadow is equal

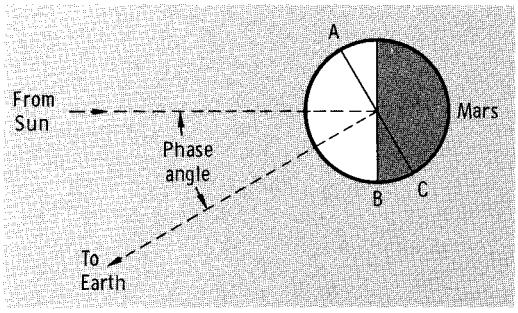


FIGURE 3.11. The phase angle and phase of Mars.

to the phase angle divided by 180 degrees. The phase angle is a maximum around the quadratures, but the actual value varies somewhat with the position of Mars in its orbit, because of the eccentricity. The maximum value, however, is 47 degrees, and the fraction of the disk in shadow is 47/180; that is, a little more than a fourth. Mars then exhibits a gibbous phase, rather like the Moon 3 or 4 days before and after its full phase (fig. 3.12). Most of the time, therefore, Mars appears in full or almost full phase.

The bright outer edge of Mars (or other

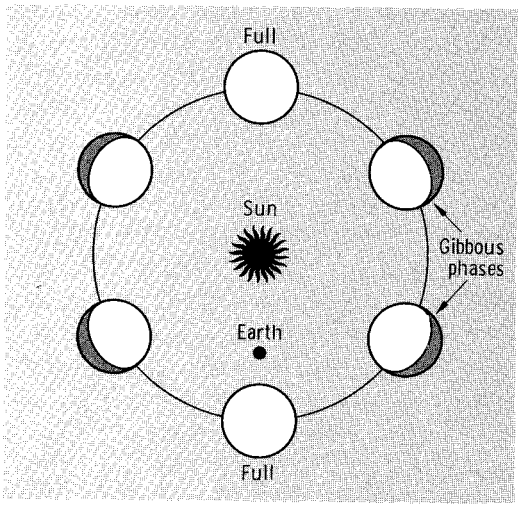


FIGURE 3.12. Gibbous and full phases of Mars.

celestial body) as seen against the darkness of space is called the limb. This would be equivalent to the point *A* in the section in figure 3.11. The edge corresponding to *B*, that is, the line separating the sunlit and shadowed areas in a gibbous phase, is referred to as the terminator, as indicated in figure 3.13.

Retrograde Motion of Mars

If a planet is observed on several successive nights, it seems to move in a general west-

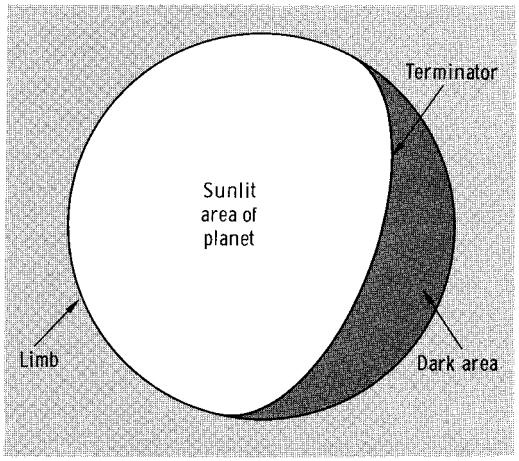


FIGURE 3.13. The limb and terminator of a celestial body.

to-east direction relative to the constellations of fixed stars. This apparent mode of movement is known as direct motion of the planet and is a consequence of Earth's orbital revolution around the Sun. The fixed stars do not exhibit such apparent motion because they are very much farther away. On certain occasions, which for Mars are invariably 4 or 5 weeks before every opposition, the direction of movement of the planet through the constellations changes direction. The planet then appears to travel from east to west. This retrograde motion, as it is called,

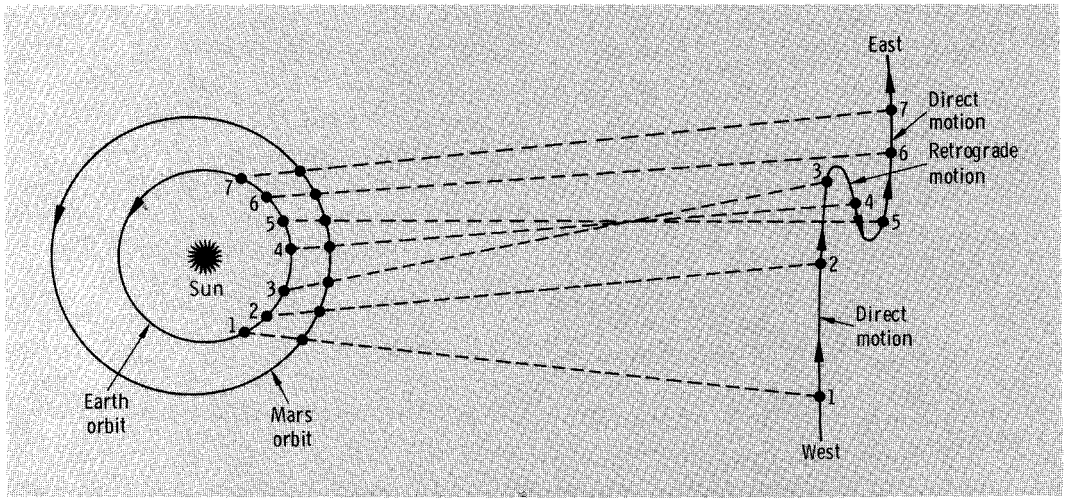


FIGURE 3.14. Explanation of direct and retrograde motion of Mars.

persists for 2 or 3 months (for Mars) before the normal direct motion is resumed.

The retrograde motion of a planet, such as Mars, as Kepler pointed out, arises from the difference in the orbital speeds of Earth and of Mars. The situation may be explained with reference to figure 3.14, which shows the simultaneous positions of Earth and Mars in their orbits around the Sun at successive times indicated by the numbers 1, 2, 3, etc. As seen from Earth, the apparent position of Mars in the sky at any time is given by the point where the line passing through the locations of Earth and the planet appears to intersect the background of fixed stars. These points, corresponding to the various simultaneous positions of the two planets, are represented by the points at the right.

An examination of the figure shows that at times 1, 2, and 3, as Earth catches up on Mars just before opposition at 4, Mars exhibits apparent direct motion, from west to east. Between times 3 and 4, as the two planets come closer together, the motion becomes retrograde, that is, from east to west, and this continues past time 4. Subsequently, Earth moves increasingly ahead of Mars and the apparent motion of Mars is again direct. It is only (and always) around the time of opposition that the conditions are suitable for apparent retrograde motion to occur. Furthermore, it can be demonstrated by means of diagrams like figure 3.14 that, at all other times, Mars exhibits direct motion among the stars, as observed from Earth.

The actual locations in the sky of Mars before and after the opposition of 1965 are represented in figure 3.15. This type of be-

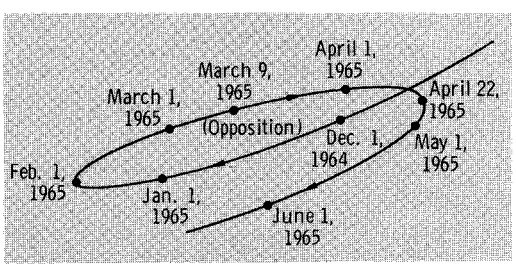


FIGURE 3.15. Apparent motion of Mars around the 1965 opposition.

havior, when the retrograde motion is higher in the sky than the direct motion, occurs at or near unfavorable (aphelion) oppositions. The movement of the planet in the loop representing the apparent path of the planet is in a clockwise direction (fig. 3.16*A*). The

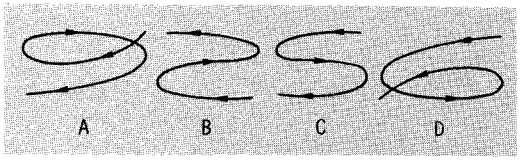


FIGURE 3.16. Types of motion of Mars near opposition.

reverse situation, with motion in a counterclockwise loop (fig. 3.16*D*), is observed at (or near) the most favorable (perihelion) oppositions. At other times, between the most and least favorable oppositions, the paths of the planet through the sky exhibit intermediate Z- or S-shapes (fig. 3.16 *B* and *C*). It is because the orbital planes of Mars and Earth do not coincide that the form of the apparent path of Mars in the sky depends on where in the Martian orbit the opposition occurs.

The Orbital Plane of Mars

The orbit of a planet lies in a plane and this plane must pass through the Sun. Although the orbital planes of the various planets are not very far apart, they do not coincide. The apparent path, from day to day, of the Sun among the constellations is referred to as the ecliptic (fig. 3.17). This name arises from the fact that all eclipses, both of the Sun and the Moon, are observed to occur, as indeed they must, along this imaginary circle. The ecliptic, or apparent path of the Sun in the sky, is inevitably an extension of the plane in which Earth orbits the Sun. Hence, Earth's orbital plane is also called the ecliptic plane.

From extensive observations on Mars, it has been concluded that the plane of its orbit around the Sun crosses the ecliptic plane at an angle of 1.85 degrees. This angle is not large, but it has some significant consequences. The illustration in figure 3.18 is an exaggerated representation of an edge-on view of the orbital planes of Earth and Mars. The orbital plane of Mars is to the south of the ecliptic plane (see fig. 3.17 for directions of north and south) near the Martian perihelion and to the north around aphelion. Because of the great distance from the Sun to Mars, the small angle of inclination of the orbital planes of Mars and Earth is sufficient to bring Mars about 6.6 million kilometers (4.1 million miles) south of the ecliptic plane at perihelion and 8.0 million kilometers (5 million miles) north of it at aphelion.

As a consequence of the inclination of the orbital planes, when Earth begins to overtake Mars shortly before a perihelic (favorable) opposition, Mars appears at first to sink increasingly farther south of the ecliptic; that is, it appears lower and lower in the sky on successive days. The reason is that the

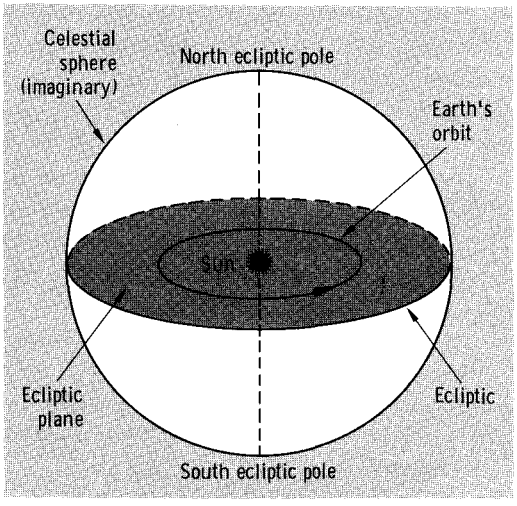


FIGURE 3.17. The ecliptic plane and poles.

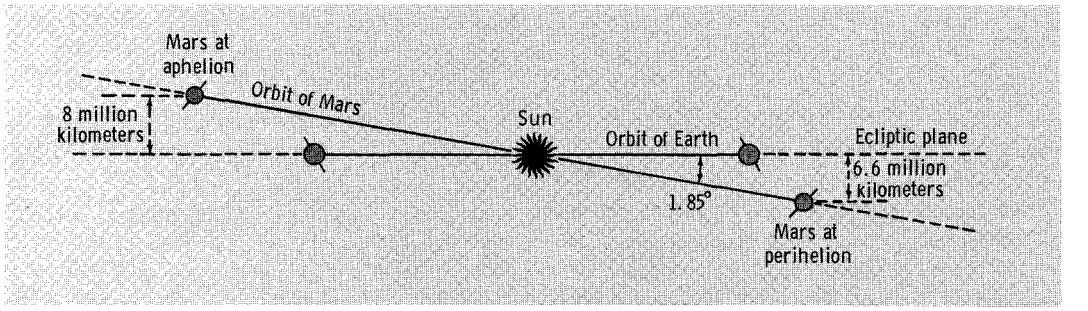


FIGURE 3.18. Relative positions (exaggerated) of the orbital planes of Earth and Mars.

distance the orbital plane of Mars is south of the ecliptic increases as perihelion is approached (fig. 3.18). After opposition, however, when Mars has moved past its perihelion, the planet appears higher in the sky from day to day, because the distance between the orbital planes is now decreasing. It is thus possible to account for the counterclockwise loop (fig. 3.16*D*) made by the apparent path of Mars at the time of a perihelic opposition.

Near an aphelic (unfavorable) opposition, the situation is reversed. As Earth approaches Mars, the latter planet moves farther and farther north of the ecliptic and it will be seen steadily higher in the sky. After opposition, Mars will cease to move higher and will start to sink. The loop formed will then have the clockwise form in figure 3.16*A*. At other (intermediate) oppositions, Mars will appear to move continuously either higher (farther north) or lower (farther south) in the sky, depending on the location of the planet in its orbit. The path will then have either a Z- or an S-shape, respectively, as in figure 3.16 *B* and *C*.

The period during each day when Mars appears above Earth's horizon, and which determines the time for which the planet might be visible, depends on the location of the observation point on Earth. The situation is related to the direction of Earth's axis of rotation, as will now be shown. Earth rotates once daily about an axis passing through the North and South Poles. But this axis is not at right angles to the orbital (or ecliptic) plane. The angle of inclination of Earth's axis to a line perpendicular to the orbital plane, as represented in figure 3.19, is close to 23.5 degrees. Incidentally, this angle is also the inclination of the equatorial plane of the planet to its orbital plane. The direction of the axis of rotation remains essentially unchanged regardless of the position of Earth in its orbit. There are actually small variations resulting from the gravitational attractions of the Sun and Moon, in particular. However, they need not be considered here.

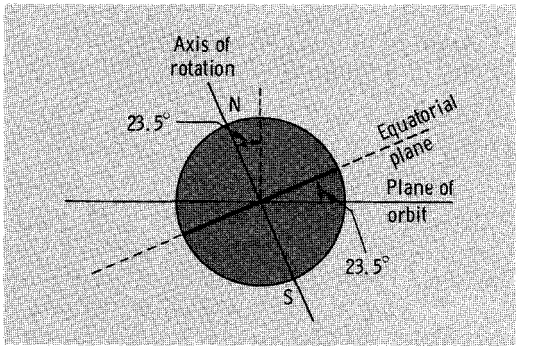


FIGURE 3.19. Inclination of Earth's axis.

Observations From Northern and Southern Hemispheres

It will be recalled that the favorable oppositions of Mars occur in July, August, and early September when it is summer in Earth's Northern Hemisphere. During this period, Earth's axis is in such a direction that the North Pole is tilted toward the Sun (fig. 3.20). Mars is close to its perihelion, which is slightly below (south of) the ecliptic plane, and is on the side of Earth opposite to the Sun. Consider a location *P* on Earth at a latitude of about 40° N, which is approximately that of Washington, D.C. The directions of the horizon and of the zenith (directly overhead) at this point are indicated in the figure.

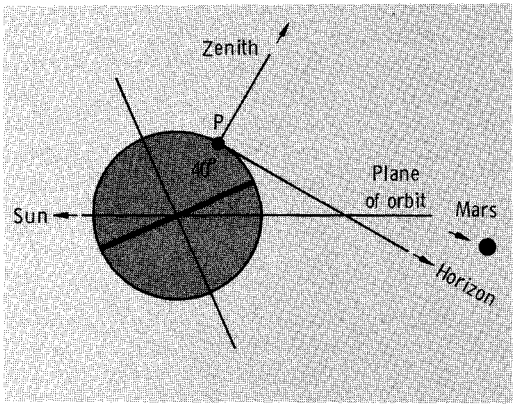


FIGURE 3.20. Observation of Mars at a favorable opposition from 40° N latitude on Earth.

An examination of figure 3.20 shows that, for the particular circumstances represented, to an observer at 40° N latitude, Mars will appear low on the horizon at the time of a favorable opposition. The same general situation applies to all points in the Northern Hemisphere, although Mars is seen somewhat higher in the sky at lower latitudes; that is, nearer to the Equator. The apparent daily (or nightly) paths of celestial bodies at different locations in the sky, as observed at a latitude of about 40° N, are indicated in figure 3.21. The lower the maximum elevation

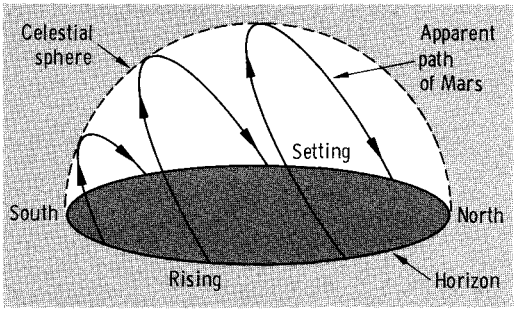


FIGURE 3.21. Paths of celestial bodies as observed at 40° N latitude.

above the horizon, the shorter the period the body, such as Mars, remains visible.

It follows, therefore, as stated earlier, that the times of favorable (perihelic) oppositions of Mars are not very good for observation from the Northern Hemisphere. The conditions improve toward the Equator and are excellent in the middle latitudes of the Southern Hemisphere where Mars rises high in the sky during the long nights of the local winter.

At aphelic (least favorable) oppositions, it is winter in Earth's Northern Hemisphere and summer in the Southern Hemisphere. The South Pole is then tilted toward the Sun (fig. 3.22). The conditions for observing Mars are then better in the Northern than in the Southern Hemisphere. However, at aphelic oppositions, Mars is much farther from Earth than at perihelic oppositions.

At intermediate oppositions, between aphelion and perihelion, especially those which occur in late May and early June or late November and early December, the conditions for observing Mars are equally good

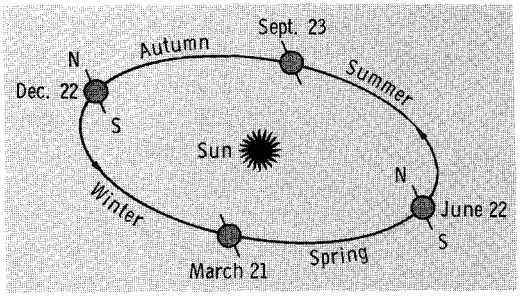


FIGURE 3.22. Tilt of Earth's axis at various times; seasons indicated are those in the Northern Hemisphere.

in both Northern and Southern Hemispheres. They are best near the Equator, where Mars is almost directly overhead at midnight at the time of opposition.

Inclination of the Axis of Mars

Mars, like Earth, also rotates about a north-south axis, and this axis is inclined at an angle of about 25 degrees to a line perpendicular to the orbital plane. Just as the inclination of Earth's axis determines the best locations from which to observe Mars, so the inclination of the Martian axis determines the parts of Mars that can be seen at the various oppositions.

Although the angles of inclination of the axes of Mars and of Earth (23.5 degrees) are not very different, the two axes point in different directions in space. The two directions, in fact, make an angle of about 45 degrees, that is, half a right angle, with each other. Earth's axis, at present, points in the north close to the star Polaris (Pole Star), but the axis of Mars is directed in the sky somewhere near the star Deneb (Alpha Cygni). Both directions change slowly in the course of time because of gravitational attractions of other members of the solar system.

The directions of the axes of Earth and of Mars at various oppositions are represented approximately in figure 3.23. Because the orbital plane of Mars is inclined at such a relatively small angle (1.85 degrees) to the ecliptic plane, it is assumed, for simplicity, that the two planes coincide. Mars is thus indicated as if it lies on the ecliptic plane; although this is not strictly correct, the general conclusions to be drawn will not be affected.

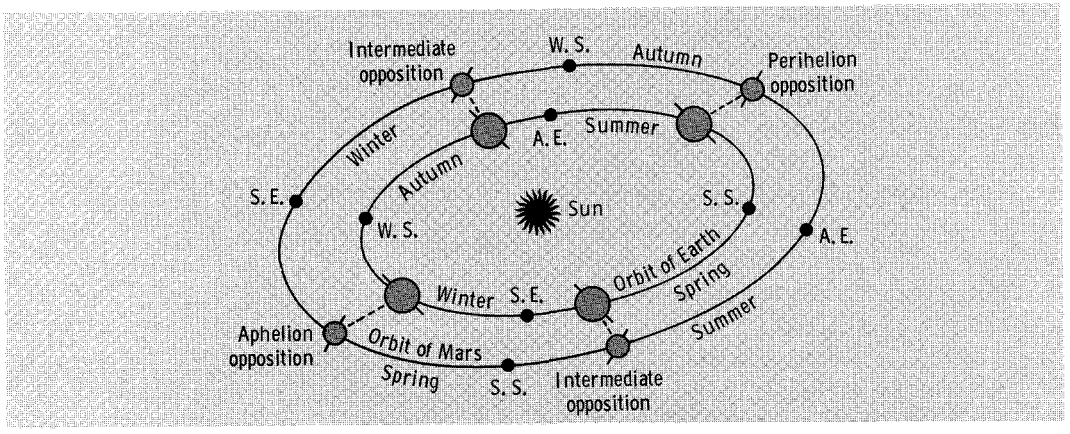


FIGURE 3.23. Direction of axes of Earth and Mars at various oppositions. (The abbreviations S.E. and A.E. stand for spring and autumn equinoxes and S.S. and W.S. for summer and winter solstices.)

It is seen from the figure that at perihelic (favorable) oppositions, when Earth's North Pole is tipped toward the Sun, the north pole of Mars is tipped away from the Sun (and Earth). Thus, at such oppositions, this pole, and the adjacent high northern latitudes, will not be visible from Earth, although the Martian south pole and the surrounding regions can be observed. Conversely, at aphelic (unfavorable) oppositions, the high northern latitudes and north pole of Mars will be seen, but the south-polar areas will be invisible. There are some differences in the parts of Mars that can be seen from different locations, such as Northern and Southern Hemispheres, on Earth. But the general situation is not affected.

When an opposition occurs approximately midway between perihelion and aphelion, at the times mentioned on page 39, it should be possible to see the whole of the Martian surface. To an observer on Earth, the axis of Mars will appear to be almost perpendicular to the orbital (or ecliptic) plane. At the time of opposition, the entire Martian surface, including both poles, should come into view as the planet rotates.

The Seasons on Earth

Just as the inclination of Earth's axis is responsible for seasonal changes in the lengths of daylight and darkness, and hence in the temperature, so the inclination of the axis of Mars is the cause of similar seasons on Mars. The circumstances can be explained most conveniently by considering, first, the situation on Earth.

As Earth revolves in its orbit about the Sun, its axis of rotation always points in the same direction in space, as indicated in figures 3.22 and 3.23. As a result of the constant direction of Earth's axis, there is one day each year, June 22, when the North Pole is tilted toward the Sun by the maximum angle of 23.5 degrees to the vertical. At the same time, the South Pole is tilted to the same extent away from the Sun. Exactly half a year later, on December 22, the situation is reversed, and the North Pole is tilted 23.5 degrees away from the Sun and the South Pole is tilted the same angle toward the Sun. At the intermediate dates of March 21 and September 23, Earth's axis is perpendicular to the line joining Earth to the Sun, and neither the North nor the South Pole is directed either toward or away from the Sun. The situation may perhaps be seen more clearly in figure 3.24, which is a representation of Earth at various points in its orbit as seen from the north ecliptic pole (fig. 3.17). The heavy dots indicate the position of Earth's North Pole.

During the period from March 21 to June 22, as the tilt of the North Pole toward the Sun increases from zero to 23.5 degrees, it is spring in the Northern Hemisphere. Then,

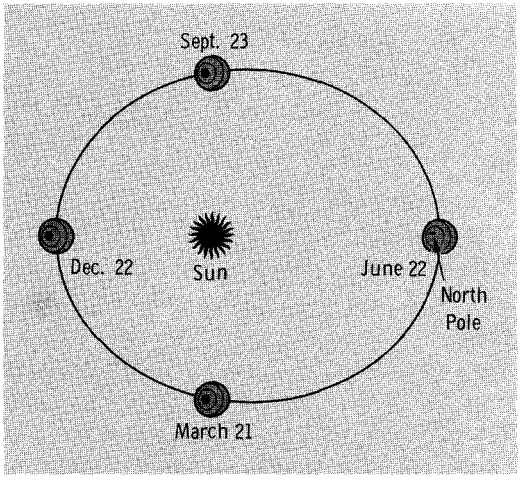


FIGURE 3.24. Earth's axis at the equinoxes and solstices as seen from the north ecliptic pole. (Eccentricity of the orbit is exaggerated for clarity.)

from June 22 to September 21, as the tilt decreases to zero again, the Northern Hemisphere experiences summer. During autumn, between September 21 and December 22, the North Pole tilts increasingly away from the Sun and, finally, between December 22 and March 21, it is winter in the Northern Hemisphere. Because the tilt of the South Pole to or away from the Sun is opposite to that of the North Pole, the seasons are reversed. Thus, spring, summer, autumn, and winter in the Northern Hemisphere correspond to autumn, winter, spring, and summer, respectively, in the Southern Hemisphere. The dates and lengths, to the nearest day, of the seasons in the two hemispheres are given below.

Dates	Northern Hemisphere	Southern Hemisphere	Days
Mar. 21 to June 22..	Spring...	Autumn..	93
June 22 to Sept. 23..	Summer..	Winter...	93
Sept. 23 to Dec. 22..	Autumn..	Spring...	90
Dec. 22 to Mar. 21..	Winter...	Summer..	89

It will be noted that spring and summer in the Northern Hemisphere are longer than autumn and winter. This is because Earth is in the vicinity of its aphelion during the former period (figs. 3.22 and 3.24) and is consequently moving more slowly in its orbit than it does near perihelion. Thus Earth is actually farther from the Sun during the summer in the Northern Hemisphere than it is in winter. In the Southern Hemisphere the situation is, of course, reversed. The summers in the Southern Hemisphere are consequently slightly shorter but hotter and the winters longer and colder than in the Northern Hemisphere.

The dates March 21, when winter turns to spring in the Northern Hemisphere, and September 23, when summer turns to autumn, are referred to as the spring (vernal) and autumnal equinoxes, respectively. The word equinox, from the Latin meaning equal night, refers to the fact that on these dates the day and night are everywhere of equal length (12 hours). The reason is that Earth's axis is then perpendicular to the Earth-Sun line, as seen earlier, and equal parts of the globe are in sunlight and darkness. The intermediate dates, June 22 and December 22, are called the summer and winter solstices, respectively. The times of rising and setting of the Sun on successive days do not change significantly at the solstices, and this has led to the use of the term which means Sun stands still.

The differences of temperature experienced in the four seasons are determined primarily by two related factors; namely, the duration of daylight and the maximum height reached by the Sun in the sky at noon. In the Northern Hemisphere, the period of daylight is longest at the summer solstice (June 22) and the Sun attains its maximum elevation in the sky. The Sun is directly overhead at noon at 23.5° N latitude, but at other latitudes it is not quite so high. Nevertheless, the Sun is highest in the sky on that day. Throughout the summer, the duration of daylight gets less and less until it is a minimum at the winter solstice (December 22) in the Northern Hemisphere. After this date, during the winter and spring, the hours of daylight increase again toward the maximum on June 22. In the Southern Hemisphere, the days are shortest on this date and longest on December 22.

The actual durations of day and night depend on the latitude. At the Equator (latitude 0°) there are always 12 hours of daylight and 12 hours of darkness. At higher and lower latitudes, the difference in length

between day and night changes with the season in accordance with the foregoing remarks. At high latitudes—in the polar regions—from latitude 90° minus 23.5° (66.5°) to 90° North or South, it is possible to have 24 hours (or more) of daylight at the time of a solstice. At 66.5° latitude, the maximum period of daylight is just 24 hours, but at still higher latitudes the periods of continuous daylight increase. At the actual poles, there are 6 months of daylight, during the local summer, and 6 months of darkness, during the local winter. As can be seen in figure 3.24, the north polar regions must be in continuous darkness during winter in the Northern Hemisphere and in continuous sunlight in the summer.

Obviously, the more hours of daylight, the greater the amount of heat radiation received from the Sun by a given area of Earth's surface. But this is not the only consideration. During the autumn and winter in the Northern Hemisphere, the Sun is lower in the sky than it is in the spring and summer. The Sun's rays thus fall more obliquely on the surface during the former seasons. As a result, a given quantity of heat radiation from the Sun is spread over a larger area (fig. 3.25), and it is consequently less effective in heating Earth's surface than when the Sun is higher in the sky, in the spring and summer.

If the duration of daylight and the elevation of the Sun were the only factors involved,

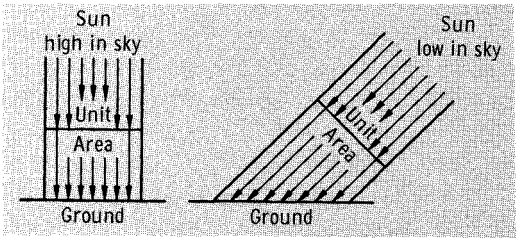


FIGURE 3.25. Radiation reaching Earth when the Sun is high (left) and low (right) in the sky.

June 22 would be the hottest day of the year and December 22 would be the coldest in the Northern Hemisphere. The actual temperatures at Earth's surface are determined by the heat balance, which is the relation between the heat lost to space and the heat gained from the Sun. The Earth tends to retain heat, especially in the oceans and the atmosphere, with the result that the highest temperatures are not reached in the Northern Hemisphere until some time after June 22. Similarly the lowest temperatures are experienced after December 22. The dates are reversed, but the delays are similar in the Southern Hemisphere.

Although the polar regions receive solar radiation continuously for periods up to 6 months, the Sun is always low in the sky. Considerable amounts of heat are absorbed during the daylight months, but much of this is expended in the partial melting of the thick layers of ice and snow. Because these layers never disappear, the surface temperatures do not increase greatly. In any event, they are always below the freezing point (0° C), although the air temperatures may be slightly higher.

The Seasons on Mars

The information given above concerning Earth can now be applied to examine the conditions on Mars. The positions of Mars in its orbit at the two Martian equinoxes and two solstices and the way Mars (and Earth for comparison) would look from the north ecliptic pole at these four locations are depicted in figure 3.26. Because of the angle between the directions in which the axis of the two planets point, it happens that the line of the equinoxes, that is, the line passing through the positions of a planet at its equinoxes, for Mars is almost at right angles to that for Earth.

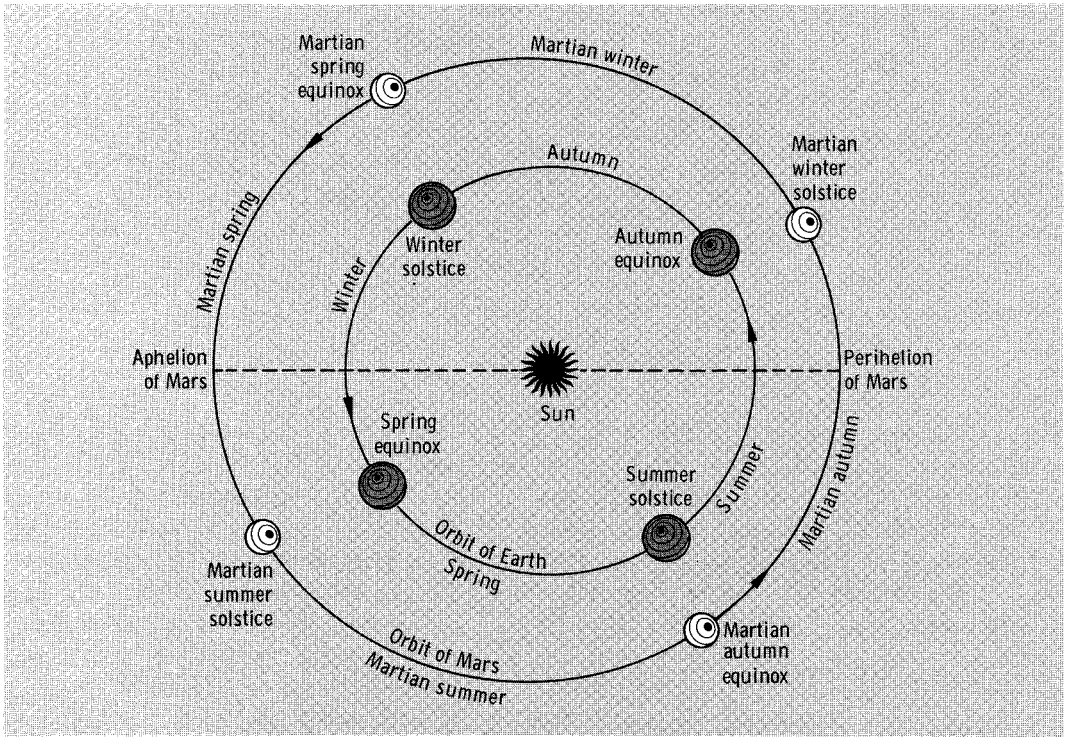


FIGURE 3.26. Seasons on Mars and on Earth.

Figure 3.26 must not be taken to imply that the spring equinox, for example, on Mars occurs within a few days of the winter solstice on Earth. What it does mean is that when Mars and Earth are close together, near an opposition, when it is December 22 on Earth, then spring will just be starting in the northern hemisphere of Mars. When a favorable opposition occurs, it is summer in the Northern Hemisphere of Earth; Mars will then be experiencing late autumn and approaching winter in its northern hemisphere. During favorable (perihelic) oppositions, the north pole of Mars is tilted away from the Sun (and Earth).

Similarly, during the unfavorable oppositions, when it is late winter in the Northern Hemisphere on Earth, it will be spring, approaching the summer solstice, in the northern hemisphere of Mars. At the intermediate oppositions in the middle and late autumn or middle and late spring on Earth, the northern hemisphere of Mars will be approaching the spring and autumn equinoxes, respectively. A rough summary of the situations for oppositions at various times of the year is given in a table on the next page.

Because there is no established Martian calendar and the Martian and terrestrial years are different, it is not possible to give any precise dates for the equinoxes and solstices on Mars. In terms of Earth dates, they vary from year to year, but they can be readily calculated if required. Of particular in-

Type of opposition	Approaching season in Earth's Northern Hemisphere	Approaching season on Mars	
		North	South
Favorable	Autumn	Winter	Summer.
Intermediate	Winter	Spring	Autumn.
	Summer	Autumn	Spring.
Unfavorable	Spring	Summer	Winter.

terest, of course, are the seasons at the times of oppositions and these can be derived from figure 3.26 or, approximately, from the table.

The lengths of the Martian seasons in terms of Earth days are, however, quite definite. They are given in the table below to the nearest whole number of days. The total is 687 days, equal to the length of the sidereal year on Mars. The durations of the seasons are seen to be roughly twice those of the seasons on Earth, because of the correspondingly greater length of the year. It will be noted, too, that the Martian seasons are of unequal durations, ranging from 146 days of the northern autumn (southern spring) to 199 days of the northern spring (southern autumn). This is a consequence of the significant eccentricity of the Martian orbit. The differences in the duration of the seasons on Earth are much less because the orbital eccentricity is relatively small.

Martian season		Earth days
Northern hemisphere	Southern hemisphere	
Spring	Autumn	199
Summer	Winter	182
Autumn	Spring	146
Winter	Summer	160

Mars is in the vicinity of the aphelion of its orbit when it is spring in the northern hemisphere. At this time, the planet is farthest from the Sun, and hence it is traveling at its lowest speed. Because Mars is then moving relatively slowly in its orbit, the time between the spring equinox and the summer solstice is exceptionally long, 199 days. Summer in the northern hemisphere is also long, 182 days, because Mars is still in the vicinity of its aphelion. Conversely, during autumn and winter in the northern hemisphere, Mars is near its perihelion and its orbital speed is near the maximum. These seasons are consequently much shorter. The situation is, of course, reversed in the southern hemisphere.

The considerably longer spring and summer in the northern hemisphere of Mars might be expected to make the summers hotter than in the southern hemisphere. This is not so, however, because of the considerable difference in the distances from the Sun at perihelion and aphelion. The ratio of the amounts of solar heat radiation received by Mars at these two locations is inversely proportional to the square of the ratio of the respective distances. The square of the ratio of the distances is $(206.66/249.22)^2$, and the inverse of this is found to be 1.44. Consequently, at perihelion Mars receives about 44 percent more radiation than at aphelion.

Orbital Characteristics of Mars

Distance, Mars to Sun (max.)	249.22×10^6 kilometers	154.86×10^6 miles
Distance, Mars to Sun (min.)	206.66×10^6 kilometers	128.41×10^6 miles
Distance, Mars to Sun (average)	228×10^6 kilometers	142×10^6 miles
Martian orbit, major axis	455.88×10^6 kilometers	283.27×10^6 miles
Martian orbit, eccentricity	0.0933	
Orbital speed (max.)	26.4 kilometers (16.4 miles) per second	
Orbital speed (min.)	22.0 kilometers (13.6 miles) per second	
Orbital speed (average)	24.1 kilometers (15.0 miles) per second	
Orbital period (sidereal)	686.980 (Earth) days	
Orbital period (synodic from Earth)	779.935 (Earth) days	
Inclination of orbit to ecliptic	1.85 degrees	
Inclination of equator to orbital plane	25 degrees	
Earth-Mars opposition distance (max.)	101×10^6 kilometers	63×10^6 miles
Earth-Mars opposition distance (min.)	56×10^6 kilometers	35×10^6 miles

During the local spring and summer in the southern hemisphere, when Mars is near perihelion, this hemisphere receives much more heat than does the northern hemisphere during the same local seasons. In spite of its shorter duration, the summer is hotter than in the northern hemisphere. As a result the south polar cap of Mars frequently disappears during the local summer, but the north polar cap has never been known to do so. Because Mars is in the vicinty of aphelion during the southern autumn and winter, the winter is colder, as well as longer, than in the northern hemisphere. The foregoing conclusions concerning the Martian seasons can thus be summarized in the following manner.

Northern Hemisphere	*Southern Hemisphere*
Winter shorter and warmer	Winter longer and colder
Summer longer and cooler	Summer shorter and hotter

The lengths of the periods of daylight and darkness of Mars during the various seasons vary in much the same way as they do on Earth. This is the case because the angle of inclination of the Martian axis is not very different from that of Earth's axis. The temperatures on Mars, however, may be expected to respond more closely to the hours of sunshine because there are no bodies of water and the atmosphere is very tenuous. Consequently, maximum and minimum temperatures should be attained soon after the local summer and winter solstices, respectively.

FOR FURTHER INFORMATION

DREYER, J. L. E.: A History of Astronomy from Thales to Kepler. Chapter XV, Dover Publications, 1953.

LEY, W.; VON BRAUN, W.; AND BONESTELL, C.: The Exploration of Mars. The Viking Press, 1956.

MICHAUX, C. M.: Handbook of the Physical Properties of the Planet Mars. NASA SP–3030, 1967.

RICHARDSON, R. S.; AND BONESTELL, C.: Mars. Harcourt, Brace & World, 1964.

IV

Mars as a Planet

PHYSICAL CHARACTERISTICS OF MARS

The Mass of Mars

THIS CHAPTER IS CONCERNED with the properties of Mars as a planet, that is, as an individual member of the solar system. The first of these properties to be considered is the mass. Two methods have been commonly used in the past to determine the mass of a planet. One is dependent on the disturbances (or perturbations) produced by the planet on the orbital motion of an adjacent planet or, preferably, of an asteroid. The other is based on a study of the orbit of a natural satellite (or moon) of the planet. In addition, a third method, which has recently become feasible, is to observe the influence of the planet on the trajectory of a spacecraft in its vicinity.

Because Mars has two satellites, and was approached by the spacecraft Mariner IV in July 1965 (ch. XIII), all three procedures outlined above have been applied in calculating the mass of the planet. Of the three, the last one is probably the most accurate and it has led to the following value of the mass of Mars:

$$6.419 \times 10^{26} \text{ grams}$$
$$\text{or } 6.419 \times 10^{20} \text{ metric tons}$$

where 1 metric ton is 1000 kilograms (2204.6 pounds). This is 0.1073, or a little over one-tenth, of the mass of Earth.

It is of interest that the mass of Mars obtained in 1877 by A. Hall from observations of the satellites of Mars, which he discovered, was 6.430×10^{26} grams. This is the value which has been used since 1895 in "The American Ephemeris and Nautical Almanac," published annually by the U.S. Naval Almanac Office. It is difficult to understand, however, the discrepancy between the Mariner IV value for the mass of Mars and that published in 1961 by D. Brouwer and G. M. Clemence, 6.4411×10^{26} grams, a figure based on some 30 observations of the satellites over a period of years. Furthermore, the mass recommended by G. M. Clemence of the Naval Almanac Office to the International Astronomical Union in 1964 was 6.521×10^{26} grams.

The Diameter of Mars

The simplest method for determining the diameter of a planet is to measure its apparent angular diameter when the planet is at a

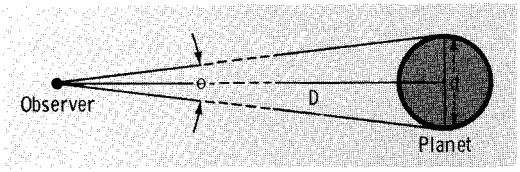


FIGURE 4.1. Determination of the diameter of Mars.

known distance from Earth. In figure 4.1, for example, in which the size of the planet is greatly exaggerated for clarity, the angular diameter is indicated by θ. If D is the distance from the planet to Earth and θ is expressed in radians (1 radian=57.296 degrees of arc), then the linear diameter d of the planet is given by

$$d=\theta\times D$$

where d and D are in the same units of length. The angle θ is determined by visual observation in a telescope or by measurement of a photographic image, and the distance D is derived from the position of the planet in its orbit at the time the angle is measured.

Because the angular diameter varies with the distance of the planet, it is the usual practice to refer the value to a standard distance. This standard distance is the astronomical unit; that is, the average distance of Earth from the Sun. The reduced angular diameter referred to the astronomical unit is equal to $\theta\times D/\mathrm{AU}$ (or d/AU), where AU, the astronomical unit, is 149.6 million kilometers (92.95 million miles). The linear diameter of the planet, expressed in astronomical units, is then equal numerically to the reduced angular diameter in radians.

The procedure described above is simple in principle, but it has proved difficult to put into practice for Mars, for reasons which will be apparent shortly. Consequently, a more indirect method has been developed for this planet. It involves observations on the apparent path, as a function of time, of a characteristic surface marking as the planet rotates.

Apart from systematic errors in measurement, arising from instrumental problems and from conditions in Earth's atmosphere, which apply to all planets, there is a peculiar difficulty in determining the angular diameter of Mars. In 1925, W. H. Wright, at the Lick Observatory in California, noted that the size of a photographic image of Mars varied with the color of the light in which the image was obtained.

A transparent blue filter allows blue light to pass through, but it will absorb, and hence stop, light of longer wavelength, such as green, yellow, orange, and red. A red filter, on the other hand, will absorb light of shorter wavelength, namely, blue, green, orange, and yellow, but will transmit red light. By the use of a series of colored filters, it is possible to obtain a photographic image of Mars in light of a particular color; that is, of a particular wavelength. Wright found that the angular diameter of Mars as determined from an image obtained in blue light was about 3 percent larger than in red light. This dependence of the apparent diameter of Mars on the wavelength of the light in which the planet is observed is called Wright's phenomenon or the Wright effect.

Although some astronomers have doubted the reality of the Wright effect, it seems that, in part at least, it is not merely caused by instrumental errors or by variations in the sensitivity of the photographic plates. A possible contributory factor is that in red light the brightness of the Martian surface appears to diminish outward from the center to the edge of the disk, a phenomenon called limb darkening (p. 44). On the other hand, in blue light, there is limb brightening, with the intensity increasing outward from the center of the disk. The limb brightening in blue

light is probably caused by the scattering (or reflection in random directions) of sunlight by the Martian atmosphere (ch. V). As a result of the limb darkening in red light and the brightening in blue light, the photographic image would appear larger in the latter case. This, however, may not be sufficient to account completely for the Wright effect.

Wright and others have suggested that the phenomenon is caused by differences in scattering of light of different wavelengths at high altitudes, for example, about 100 kilometers (62 miles), in the Martian atmosphere. But the density of the gas molecules at such an altitude is now thought to be too small to account for the Wright effect. Another suggestion is that a layer of fine dust particles high in the atmosphere causes the preferential scattering of blue light. Such a layer, which does not appear to be too probable, has been postulated to explain the blue-haze phenomenon described more fully in chapter VII. Recently, the possibility has been mentioned that the Wright effect is not a Martian phenomenon, but arises in part from conditions in Earth's atmosphere (p. 150).

No matter whether the Wright effect is real or illusory, and regardless of its cause, the fact remains that photographic (and visual) images seem to give different diameters for Mars in light of different wavelengths. Which wavelength, if any, then gives the true diameter? It has been generally accepted that images of Mars in red light give the best approximation to the actual size of the planet. Even if this were so, however, there is still a subjective problem in interpreting the telescopic images; it is difficult to determine, either by the eye or even with instruments, the precise boundary of the small bright disk of Mars. Consequently, the estimation of the angular diameter is bound to involve an element of uncertainty.

The method of determining the diameter of Mars based on measurement of the displacement of a characteristic feature during the rotation of the planet would seem to offer the prospect of improved accuracy, at least in principle. This turns out not to be the case in practice, however. There are very few, if any, sharply distinguishable markings on the surface the positions of which can be observed precisely enough to make possible an accurate calculation of the radius of Mars.

Like Earth, the planet Mars is not exactly spherical, but is somewhat flattened at the poles. In other words, the polar diameter is less than the equatorial diameter. This is shown in the section in figure 4.2 which is not drawn to scale; r_p is the polar radius, that is, half the polar diameter, and r_e is the equatorial radius, or half the equatorial diameter. Provided its interior is not completely rigid, any body, like Earth or Mars, which rotates at a relatively high speed is expected to have a larger diameter in the equatorial direction because of the centrifugal effect of the rotation. The force in the outward direction caused by this effect is a maximum at the equator of a rotating mass. Because the equatorial and polar diameters of Mars are different, they must, of course, be measured

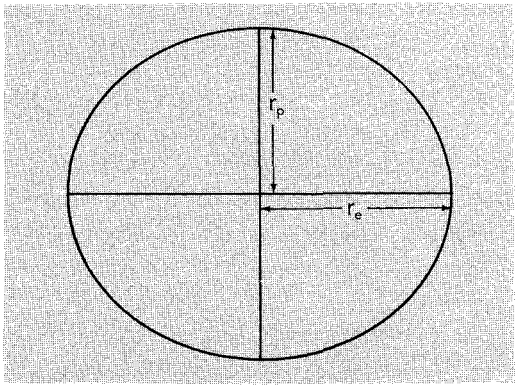


FIGURE 4.2. Polar and equatorial radii of a planet.

separately. The magnitude and significance of the difference will be discussed later in the chapter, but the results of the measurements will be given here.

In 1964, G. de Vaucouleurs in the United States published an extensive critical review of the results of determinations of the equatorial and polar diameters of Mars dating back to 1879. He concluded that the equatorial angular diameter of Mars, adjusted to an Earth-Mars distance of 1 AU, was between 9.285 and 9.415 seconds of arc. Because 1 degree of arc contains 3600 seconds, 1 radian is equivalent to $57.296 \times 3600 = 206\,265$ seconds. The linear equatorial diameter of Mars in astronomical units is then obtained upon dividing the angular diameter in seconds by 206 265. To derive the diameter in kilometers, the result is multiplied by 149.6 million.

The equatorial diameter of Mars is thus found to lie between 6723 and 6828 kilometers. By taking the various factors into consideration, de Vaucouleurs suggested a rounded figure of 6750 kilometers (or 4190 miles) for the equatorial diameter. In the same way, a value of 6700 kilometers (4160 miles) was suggested for the polar diameter. The mean diameter of the planet is thus approximately 6720 kilometers (4170 miles) and the average radius is about 3360 kilometers.

The trajectory of the Mariner IV spacecraft in the vicinity of Mars provided another means for obtaining an indication of the radius of the planet. By timing the radio signal from Mariner IV just before the spacecraft went behind the planet and again when it reappeared on the other side (fig. 5.6), it was possible to calculate the distance of the closest approach of the radio beam to the center of gravity (or center of mass) of the planet. The values found were 3384 kilometers for the distance when the radio signal was cut off (at a latitude of 60° N) and 3379 kilometers when it reappeared again (at a latitude of 50° S). The corresponding diameter, which should be less than the equatorial diameter, would then be about 6760 kilometers.

It appears, therefore, that the values of the equatorial and polar diameters suggested by de Vaucouleurs may be too low, and that a more reliable value for the average radius of Mars is approximately 3380 kilometers (2100 miles). For purposes of comparison, it may be noted that the average radius of Earth is 6371 kilometers (3959 miles). Consequently, the radius of Mars is a little more than half that of Earth. The volume of Mars, which depends on the cube of its radius, is thus approximately one-seventh the volume of Earth.

The Flattening of Mars

The flattening, also sometimes referred to as the ellipticity or oblateness, of a planet is defined by

$$f = \frac{r_e - r_p}{r_e}$$

where r_e and r_p are the equatorial and polar radii, respectively. Although the radii of Mars determined by individual observers have varied significantly, as seen above, the flattenings derived from the measurements have been in good agreement. The average value obtained from studies of photographic and visual telescopic images of the planet is given as approximately 0.011. This is called the optical flattening because it is determined from optical measurements. A flattening of this magnitude would mean that the equatorial radius of Mars is 36 kilometers (22 miles) greater than the polar radius.

There is another method for evaluating the flattening of a planet, depending on the

motion of a satellite (or moon) in an orbit that is not too far distant from the parent planet. For Mars, the required conditions are satisfied by its two satellites. The orbits of these satellites are not quite in the same plane as the Martian equator, but each orbit crosses the equatorial plane at two points, called the nodes of the orbit. The line, lying in the equatorial plane, that joins the two nodes is referred to as the line of the nodes.

If the planet were a uniform sphere, the line of the nodes of a satellite's orbit would remain stationary in space. But if the planet is flattened, the orbit of the satellite is continuously disturbed in such a way that the line of the nodes rotates in the equatorial plane. From the rate of rotation, the flattening of the planet can be calculated. This is called the dynamical flattening.

From a study of the orbits of the Martian satellites, the dynamical flattening of the planet has been calculated to be 0.00525, which is about half the optical flattening. There has been much discussion, but no conclusion, concerning this discrepancy between the dynamical and optical flattening of Mars. Because the telescopic image may depend on the possible effects arising from the scattering of light in the Martian atmosphere, it was thought that the results of the optical method might be incorrect.

In answer to this, however, reference has been made to the value reported in 1927 by R. J. Trumpler, based on observations at the Lick Observatory during the 1924 apparition. By tracking a conspicuous feature on the surface of Mars as the planet rotated, Trumpler found the flattening to be 0.011. This result, which is essentially independent of scattering in the Martian atmosphere, is the same as the optical flattening obtained by direct measurement of the equatorial and polar diameters (or radii). But the agreement is not altogether convincing because the indirect method used by Trumpler is not considered to be very accurate.

There is a possibility, however, that both the optical and dynamical flattening values may be correct. The optical flattening is determined from the actual (linear) dimensions of the planet, but the dynamical flattening depends on the mass distribution. If the mass is not distributed uniformly throughout the interior of the planet, then the optical and dynamical flattenings may not be the same. It is of interest to note in this connection that the flattening of Earth as determined from direct measurements on the surface is essentially identical with the dynamic flattening calculated from the orbital motions of artificial satellites.

The American scientist H. C. Urey suggested in 1950 that the difference between the optical and dynamic flattening of Mars could be explained if there were a belt of mountains around the equator, accompanied by a compensation of subsurface material of lower density than average. This compensation effect, known as isostasy, leading to a uniform pressure below a certain depth in the interior, is believed to occur on Earth. In the situation postulated by Urey, the mass distribution, as it affected the motion of a satellite, would not correspond with the actual linear dimensions of the planet. The optical flattening would then be larger than the dynamical value.

In order to account for the actual difference between the two values of the flattening, the elevation of the high land around the equator would have to be about 15 kilometers (49 000 feet). Consequently, in 1952, Urey withdrew his original suggestion because, as he said, "such high plateaus or mountains . . . even in tropical regions should be covered by snow as on Earth, and . . . [this is] not observed."

It will be seen in chapter VI, however, that elevation differences on the order of 10

to 15 kilometers on Mars are not improbable. And, furthermore, the ground temperature variations with altitude, which lead to the permanent snow cover on high terrestrial mountains, may be quite different on Mars. Thus, in 1966, C. Sagan and J. B. Pollack of the Smithsonian Institution Astrophysical Observatory, in Cambridge, Mass., stated that in their view: "While the entire effect [i.e., the difference between dynamic and optical flattening] may not be due to equatorial elevations . . . we believe that Urey's explanation . . . now has some measure of observational support."

A theory, similar in principle but different in detail to that of Urey, was proposed in the United States in 1962 by D. L. Lamar. He suggested that the elevation of the Martian surface increases gradually from the poles to the equator, and that this increase is compensated isostatically by adjustment in the thickness of the planet's crust; that is, in its outermost layer (p. 67). In Lamar's opinion there is an equatorial bulge, but not necessarily high mountains or plateaus in the equatorial region. Because of the isostatic compensation, the dynamic ellipticity would be less than the optical value. But no explanation has been offered for the existence of such an isostatically compensated equatorial bulge on Mars.

Apart from the suggestion, which cannot be ruled out, that the flattening discrepancy is caused by observational errors in the optical data, an elevated equatorial belt of some kind seems to be the only satisfactory way out of the dilemma. It would appear, however, that a complete solution of the problem must await the time when more reliable information concerning both the linear dimensions and mass distribution of Mars can be obtained from spacecraft passing close by the planet and from artificial satellites that orbit around it.

Density and Gravity of Mars

Two characteristics of a planet, namely, its average density and the force of gravity at the surface, are related to the mass and the radius. Because the mass of Mars is about one-tenth and the volume one-seventh of the corresponding properties of Earth, it is evident that the average density (mass per unit volume) of Mars must be less than Earth's density. The actual average density of Mars, expressed in grams per cubic centimeter (g/cm^3), can be calculated from the mass and radius given above. The volume is equal to $\frac{4}{3}\pi r^3$, where r is the average radius of the planet in centimeters (3380×10^5). The volume of Mars is found to be 1.62×10^{26} cubic centimeters. If the mass is taken to be 6.42×10^{26} grams, then

$$\text{Density of Mars}=\frac{6.42\times10^{26}}{1.62\times10^{26}}=3.96\ g/cm^3$$

Because neither the mass nor the radius of Mars is known with any degree of accuracy, it is evident that the calculated density is equally uncertain. Values ranging from 3.85 to 4.25 g/cm^3 have, in fact, been reported. All that can be said at present, therefore, is that the average density of Mars is in the vicinity of 4 g/cm^3. For comparison it may be noted that the average density of Earth is 5.52 g/cm^3.

The gravitational acceleration at a specified point on the surface of a planet is given by the expression

$$\text{Acceleration}=\frac{G\times M}{r^2}$$

where G is the universal constant of gravitation (6.67×10^{-8} in the centimeter-gram-second system of units), M is the mass of the planet, and r is the radius of the planet at the point under consideration. The radii of

flattened bodies, like Mars and Earth, vary with the latitude, and so, consequently, also do the gravitational accelerations. For the present purpose, however, it is sufficient to determine an average value for Mars. Thus, by using the mass (in grams) and average radius (in centimeters) given above, it is found that

$$\text{Gravitational acceleration} = \frac{(6.67\times10^{-8})(6.42\times10^{26})}{(3.38\times10^{8})^2} = 375 \text{ cm/sec/sec.}$$

Like the density, the gravitational acceleration at the surface of Mars cannot be determined precisely because of the uncertainties in the mass and radius involved in the calculation. Published values range from about 360 to 390, and an approximate mean is 375 cm/sec/sec for the average acceleration due to gravity at the Martian surface. The acceleration is somewhat greater at higher latitudes, that is, nearer to the poles, and smaller at lower latitudes, toward the equator. The average value of the gravitational acceleration at Earth's surface is 983 cm/sec/sec. Hence, the gravitational force at the surface of Mars is 0.38 (almost two-fifths) of the force on the same mass on Earth.

The Martian Day

The length of a day depends on how the day is defined; for practical purposes, the sidereal day and the solar day may be distinguished. The sidereal day of a planet is the exact time required for the planet to make a complete rotation about its axis; it is also referred to as the period of rotation of the planet. For Mars, this period is determined by timing the passage of conspicuous markings on the planet past a distant fixed star. The accepted value of the period of rotation (or sidereal day) for Mars is 24 hours 37 minutes 22.67 seconds. The hours, minutes, and seconds refer to time intervals on Earth, which are defined in the manner explained below.

Timekeeping on Earth is based on the solar day, rather than on the sidereal day. Because Earth cannot be continuously observed from a great distance, as Mars is, the period of rotation must be obtained in a different manner. It is equal to the time interval between two consecutive transits (or crossings) by a distant fixed star of the meridian at any given point on Earth. The meridian is an imaginary circle in the sky that passes directly overhead and extends in a north-south direction.

The solar day, on the other hand, is the interval between two consecutive meridian transits of the Sun; that is to say, it is the interval between two consecutive noons. The solar day is longer than the sidereal day because the planet moves in an orbit around the Sun. The difference may be explained with the aid of figure 4.3. An observer at position *a* in figure 4.3, I–A, sees a distant fixed star

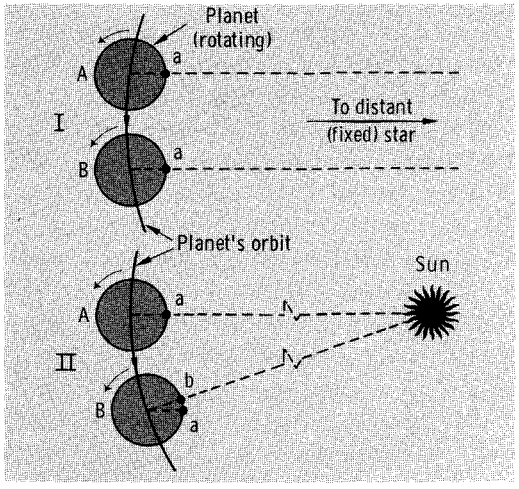


FIGURE 4.3. Comparison of sidereal (I) and solar (II) days.

cross the meridian, and exactly one sidereal day later Earth has rotated to position I–B (and has moved in its orbit) when the fixed star is again seen to cross the meridian. In figure 4.3, II–A, the Sun makes a transit of the meridian at *a,* and the next transit, a solar day later, is indicated by *b* in figure 4.3, II–B. In order to reach the point *b* where the transit of the Sun occurs, Earth will have rotated the additional amount *ab*. Consequently, the solar day on Earth is longer than the sidereal day. The same is true for Mars, although the difference in length is not the same.

It can be seen from figure 4.3 that the reason the sidereal and solar days for a given planet are different is that the planet has moved in its orbit around the Sun. If the planet simply rotated in a stationary position, the two kinds of day would be the same. The difference in the lengths depends on the distance the planet travels in its orbit between two successive noons. Since the average orbital velocity of Mars (24.1 km/sec) is less than that of Earth (29.8 km/sec), but the days are approximately the same length, Mars moves a smaller distance in its orbit than does Earth in this period. The difference between the sidereal and solar days on Mars, therefore, is somewhat less than on Earth.

The length of a sidereal day is essentially constant, but the solar day is not. It varies with the position of the planet in its orbit. The average of the lengths of the solar day throughout the year is called the mean solar day, and the mean solar day on Earth is defined as exactly 24 hours, with each hour having 60 minutes and each minute 60 seconds. The mean solar day is 3 minutes 55.91 seconds longer than the sidereal day on Earth; the latter is 23 hours 56 minutes 4.09 seconds.

On Mars, the solar day, the period between two successive transits of the meridian by the Sun, is 2 minutes 12.56 seconds longer than the sidereal day of 24 hours 37 minutes 22.67 seconds. The length of the solar day on Mars is consequently 24 hours 39 minutes 35.23 seconds. The average conventional day on Mars, the time interval between one noon and the next, is thus only about 40 minutes longer than the average terrestrial (solar) day.

From the known approximate equatorial radius of Mars and the period of rotation, the speed of rotation at the equator can be calculated. It is equal to 2π times the radius, that is, the circumference of the planet at the equator, divided by the period of rotation. The value obtained in this manner is 0.26 km/sec (0.16 mps), compared with 0.47 km/sec (0.29 mps) for Earth.

Magnetic Field of Mars

An attempt to measure the magnetic field of Mars was made by means of instruments on Mariner IV. But even at its closest approach of 9846 kilometers (6118 miles) to the planet, the magnetometer on the spacecraft detected no increase in the magnetic field strength from the value in interplanetary space. From this observation it has been concluded that the strength of the Martian magnetic field cannot be more than a very small fraction (less than 0.03 percent) of the magnetic field of Earth.

Confirmation of the extremely small magnetic field of Mars was obtained from the failure of instruments on Mariner IV to detect trapped charged particles in the vicinity of the planet. Earth has an extensive radiation (Van Allen) belt in which electrically charged particles, namely positively charged hydrogen ions (protons) and negatively charged electrons, originating in the Sun, are trapped and confined by the magnetic field. The same charged particles, although in somewhat

smaller numbers, are undoubtedly available near Mars, and the absence of a radiation belt can be ascribed only to the very small (or zero) magnetic field.

According to current views, there are two requirements for a planet to have a significant magnetic field. One is that the planet should rotate fairly rapidly, and the other is that it should have a central core of a liquid that is able to conduct electricity. Earth's core is believed to consist mainly of the heavy metals, iron and nickel. Mars satisfies one of the requirements, because it rotates quite rapidly. The absence of an appreciable magnetic field would imply that it either does not have a liquid-metal core or that such a core, if it is present, must be small in proportion to the size of the planet. The problem of the Martian core will be taken up shortly in the section dealing with the internal structure of the planet.

The Albedo of Mars

The albedo, from the Latin *albus* meaning white, is a measure of the ability of a body or material to reflect light. There are various ways of defining the albedo of a planet, and the one in common use is that suggested by W. C. Bond in the United States in 1861 and adopted by H. N. Russell in 1916. It is known as the Bond (or Russell-Bond) albedo, or as the spherical albedo. It is defined as the fraction of the solar radiation (of a given wavelength) falling on a planet and being reflected in all directions. The Bond albedo A can be divided into two factors, the geometric albedo, p, and the phase integral, q, so that

$$A = pq$$

The geometric albedo is a measure of the sunlight (of a particular wavelength) that is reflected by the planet in the direction of an observer on Earth. It is determined from the measured brightness (or luminance) of the planet at full phase, referred to standard distances from the Sun and Earth. The phase integral can be calculated, in principle, from the observed variation of the brightness with the phase angle (p. 43). For Mars, this variation is relatively small and so the phase integral is calculated on the basis of certain assumptions.

The dependence of the Bond albedo, obtained in this manner, on the wavelength (or color) of the light is shown in figure 4.4. The albedo of Mars is seen to increase from quite small values (0.05 or less) in the ultraviolet and blue regions of the spectrum to more than 0.3 in the red and infrared. It is this increase which accounts for the reddish appearance of the planet. The Bond albedo of Mars in visible light is taken to be about 0.17. The foregoing values are averages which apply to the planet as a whole. The albedos of the bright areas are higher and those of the dark areas are lower than average (ch. VI).

Another important quantity is the radiometric (or integral) albedo. This is the fraction of the total solar radiation energy that is

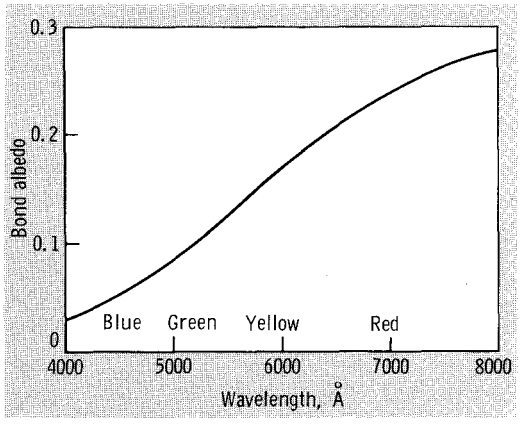


FIGURE 4.4. Dependence of the Bond albedo of Mars on the wavelength.

reflected by the planet in all directions. If the radiometric albedo is represented by A^*, then $1-A^*$ is the fraction of the Sun's radiation absorbed by the planet. The determination of A^* is hampered by a lack of observational data, but the value of 0.295, suggested by G. de Vaucouleurs in 1964, was accepted for a time. A more recent analysis, made by R. G. Walker and C. Sagan in 1968, however, indicates that A^* is closer to 0.23, and this will be adopted here. The fraction of the solar energy falling on Mars and being absorbed by the planet is thus $1-0.23$, that is, 0.77.

Energy from the Sun reaches Earth at the rate of 2.0 cal/cm²/min. As Mars is farther from the Sun, the solar radiation energy is somewhat less. The actual rate, which is inversely proportional to the square of the distance from the Sun (p. 54), ranges from 0.72, when Mars is farthest from the Sun, to 1.05 cal/cm²/min, when Mars is closest to the Sun. The mean value over the whole of the Martian orbit is 0.86 cal/cm²/min. Because only the fraction 0.77 is absorbed, the average rate at which solar energy is taken up by the Martian surface is $0.86\times0.77=0.66$ cal/cm²/min.

The Interior of Mars

There has been considerable speculation and many differences of opinion concerning the nature of the interior of Mars. All that can be said definitely at the present time is based on the absence of a significant magnetic field. This implies that Mars either has no core of liquid heavy metals, such as Earth has, or if it does have a liquid core it must be relatively small. Because the average density of Mars is substantially less than that of Earth, it might be thought that Mars has a smaller proportion of the heavier (high-density) metallic elements, such as iron and nickel, which would be present in a core. This is not necessarily true. If the heavy elements were in the form of oxides, rather than in the metallic state, the average density of Mars would be lower, even though the overall elementary composition (including the atmosphere) were the same as it is for Earth.

In 1743, the French mathematician A. C. Clairaut considered the properties of a rotating, compressible fluid body. He made use of a quantity, commonly represented by ϕ, which is the ratio of the centrifugal force at the equator of the rotating body to the force of gravity at the same location. Numerically, ϕ is found to be equal to $3\pi/GP\rho$, where G is the universal constant of gravitation, given on page 62, P is the sidereal period of rotation, and ρ is the average density of the body. Clairaut showed that for a rotating homogeneous body, that is, one with a uniform distribution of mass throughout its interior, the ratio f/ϕ, where f is the dynamical flattening, should be 1.25. On the other hand, if all the mass of the body is concentrated at its center, then f/ϕ should be 0.5.

In applying the foregoing criterion to Mars, use of 0.00525 for the dynamical flattening makes f/ϕ equal to about 1.15. As the average density is somewhat uncertain, so also is the value of the ratio f/ϕ. For comparison, the value of f/ϕ for Earth is 0.97. Hence, it would appear that the interior of Mars is much more uniform than Earth's interior, because 1.15 is closer than is 0.97 to the ideal value of 1.25 for a completely uniform body.

If the optical flattening were used in determining f/ϕ, the value would be 2.3; this would imply that the density of Mars is much less at the center than in the outer layers. Such a highly improbable situation suggests that the actual surface of Mars, as implied by the optical flattening, is not indicative of the mass distribution in the interior. Lamar has

claimed, however, that if allowance is made for the equatorial bulge which he postulated (p. 62), the optical flattening leads to a value of between 1.05 and 1.09 for the f/ϕ ratio.

A considerable amount of information about the interior of Earth has been inferred from the velocity of propagation of seismic waves and from oscillations accompanying earthquakes. The general view at the present time is that there are three main concentric regions in Earth's interior; they are, first, a very dense central core, then a less dense mantle, and finally a relatively thin crust (fig. 4.5). The core is probably divided into a solid inner zone and a liquid outer one. Both of these zones are believed to consist mainly of iron with about 10 percent nickel. Motions within the liquid part of the core are thought to be responsible for the terrestrial magnetic field. Both the mantle and the crust are probably made up of silicates of iron, magnesium, and aluminium. In the crust, which varies in thickness from about 5 to 50 kilometers (3 to 30 miles), aluminium predominates, but the exact composition of the mantle, which extends to a depth of about 2900 kilometers (1800 miles), is uncertain.

Most authorities agree that Earth and the other planets, at least those, like Mars, which are similar to Earth, were formed by the accretion of relatively cold (or moderately warm) particles. In the course of time, energy released by the radioactive decay of uranium, thorium, and potassium-40, in particular, resulted in a considerable increase of temperature. In Earth's interior the temperature rose high enough to permit iron and nickel to melt; the heavy liquid then sank and collected near the center of the planet to form the core, part of which later solidified under the existing high pressure. As a result of this process of differentiation, as it is called, the proportion of iron in the mantle and crust are less than the average in Earth's interior.

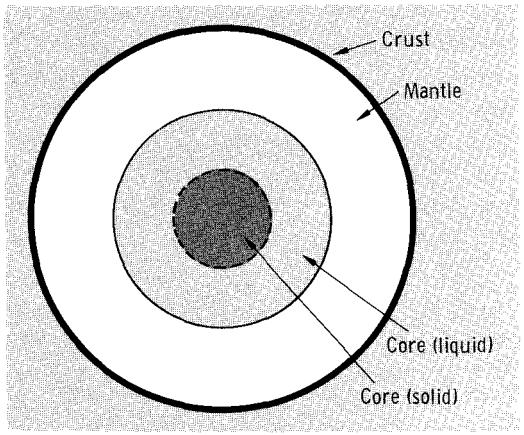


FIGURE 4.5. Section through Earth's interior showing different regions.

By analogy with Earth, it is assumed that Mars also consists of concentric regions, but it is not certain whether there are two or three. Thermal calculations have been made based on certain assumptions (postulates) concerning the initial temperature of the planet when it was first formed, the amounts and distribution of radioactive elements, and the methods of heat transfer from one point to another in the interior. Some scientists find that the temperatures attained were never high enough to permit iron to melt, so that Mars could not have a central metal core, as Earth does. Others, using slightly different postulates, claim that the formation of a liquid metal core is possible. The insignificant magnetic field of the planet suggests, however, that if such a core exists it must be small.

There are indications, as will be seen in chapter VI, that the surface of Mars may contain considerable quantities of iron oxides, with a normal density approaching 4 g/cm^3. Although this is close to the average density of the whole planet, it must be remembered that much of the interior is probably at a high temperature as well as at a high pressure; the average density of the planetary ma-

319-889 O—69——3

terial under normal conditions might thus be significantly higher. Nevertheless, it would seem that the outer layers of Mars may contain a larger proportion of iron than do the crust and mantle of Earth. This would suggest that differentiation, leading to a liquid metal core, may not have occurred on Mars, at least not to any great extent. There is a possibility, however, as will be seen on page 111, that the high iron content of the Martian surface may be the result of the deposition of meteorites.

The value of the Clairaut ratio f/ϕ for Mars (about 1.1) is less than that expected for a completely homogeneous (uniform) body (1.25), but is greater than f/ϕ for Earth (0.97). It would seem reasonable to assume, therefore, that most, if not all, of the interior of Mars consists of two concentric regions. These have been compared to the crust and mantle of Earth, although the Martian crust might be somewhat thicker than the terrestrial one.

Even if the temperatures in the interior of Mars have never been as high as the melting point of iron (from 1500° to about 2000° C at the pressures in the interior of Mars), the conditions of temperature and pressure would be adequate to permit the formation of various silicates. The separation into two regions might then be caused either by differences in chemical composition (and density) or to a change in phase; that is, in the crystalline form (and density) at a particular temperature and pressure. The more dense region would form the equivalent of the mantle, whereas the less dense one would constitute the outer layers (or crust) of Mars.

It must be admitted that relatively little is known about the structure of the interior of Mars. Furthermore, calculations are hampered by uncertainties in the average density, the radii, the dynamical flattening, and the composition. It will probably be necessary to wait until seismic instruments can be landed on the surface and spacecraft can be placed in orbit around the planet before more detailed knowledge will be available concerning the Martian interior.

THE SATELLITES OF MARS

Predictions of Martian Moons

In the year 1610, J. Kepler predicted that Mars had two satellites. This prediction turned out to be correct, although its basis was highly fortuitous. One of the earliest observations made by Galileo with his telescope was that Jupiter had 4 moons, although it is now known to have 12. Furthermore, it will be recalled from chapter III that Kepler had suspected there might be a missing planet between Mars and Jupiter. Consequently, he suggested the following simple sequence for the successive planets: Venus (no moons), Earth (one moon), Mars (two moons), missing planet (three moons), and Jupiter (four moons). What Kepler would have done had Galileo been able to detect all 12 satellites of Jupiter cannot be imagined.

The idea that Mars might have two satellites was introduced by the British writer Jonathan Swift in his famous satirical work "Gulliver's Travels," published in 1726. There is no doubt, from internal evidence, that Swift was familiar with Kepler's third law relating the period of a satellite to the dimensions of its orbit. Consequently, it is probable that he also knew of the prediction concerning the moons of Mars and adopted it in his description of Lemuel Gulliver's visit to the flying island of Laputa. The inhabitants of Laputa, wrote Swift, "have . . . discovered two lesser stars, or satellites, which revolve about Mars." The idea that Mars has two moons was also mentioned by Voltaire, apparently under the influence of Swift, in his novel "Micromégas," published in 1750.

Discovery of the Satellites

An unsuccessful search for the satellites of Mars was made by William Herschel in 1783 and his failure appears to have convinced astronomers that there were none. Nevertheless, during the 19th century, astronomers continued to look for Martian moons. One of the most careful searches was that of H. L. d'Arrest at the Copenhagen Observatory in 1864, but he was no more successful than his predecessors. The actual discovery of the two satellites of Mars was finally made in 1877 by Asaph Hall at the U.S. Naval Observatory in Washington, D.C.

"In the spring of 1877," wrote Hall, "the approaching favorable opposition of the planet Mars attracted my attention, and the idea occurred to me of making a careful search . . . for a satellite of this planet." The search was commenced in early August, and "on the night of the 11th, and at half past two o'clock," said Hall, "I found a faint object . . . a little north of the planet, which afterward proved to be the outer satellite." Cloudy weather and poor seeing conditions interfered with the observations on subsequent days, but "on August 16, the object was again found . . . and the observations . . . showed that it was moving with the planet. . . . On August 17, while waiting and watching for the outer satellite, I discovered the inner one. The observations of the 17th and 18th put beyond doubt the character of these objects."

In naming the moons of Mars which he had discovered, Hall stated:

> Of the various names that have been proposed for these satellites, I have chosen those suggested by Mr. Madan of Eton, England, viz: Deimos for the outer satellite; Phobos for the inner satellite. These are generally the names of the horses that draw the chariot of Mars; but in the lines referred to [in the Fifteenth Book of the Iliad] they are personified by Homer, and mean the attendants, or sons of Mars. [According to Bryant's translation] "He [Ares, i.e., Mars] spake, and summoned Fear [Phobos] and Flight [Deimos] to yoke his steeds."

Characteristics of the Satellites of Mars

The moons of Mars differ from Earth's Moon in several respects, and this has led to the belief that they originated in a different manner. The Martian satellites are very small in size and their orbits are close to the parent planet. The planet and its satellites are shown in the photograph in figure 4.6. The positions are correct but the images of the satellites are larger than the actual sizes. Because of the small dimensions of the orbits, which are almost circular, the periods of revolution are quite short. Furthermore, the planes of the orbits of the Martian satellites are very close to the plane of the planet's equator. For Earth's Moon, however, there is an inclination of nearly 7 degrees between the orbital plane and Earth's equatorial plane.

The Martian satellites are too small for their sizes to be determined directly from measurements of the angular diameters. Highly approximate values have, therefore, been estimated from the brightness and an assumed albedo. In this manner, the diameter of Phobos, the inner moon, is found to be roughly 16 kilometers (10 miles) and that of Deimos, the outer moon, is only about 8 kilometers (5 miles).

The orbit of the inner satellite, Phobos, has a small eccentricity, namely 0.02, so that it is almost circular. The average radius of the orbit is 9450 kilometers (5850 miles). Phobos, therefore, is only about 6100 kilometers (3750 miles) from the surface of Mars. These distances may be compared with the average radius of 384 000 kilometers (239 000 miles) of the Moon's orbit around Earth.

It is of interest that the radius of the orbit of Phobos is about 2.8 times the radius of

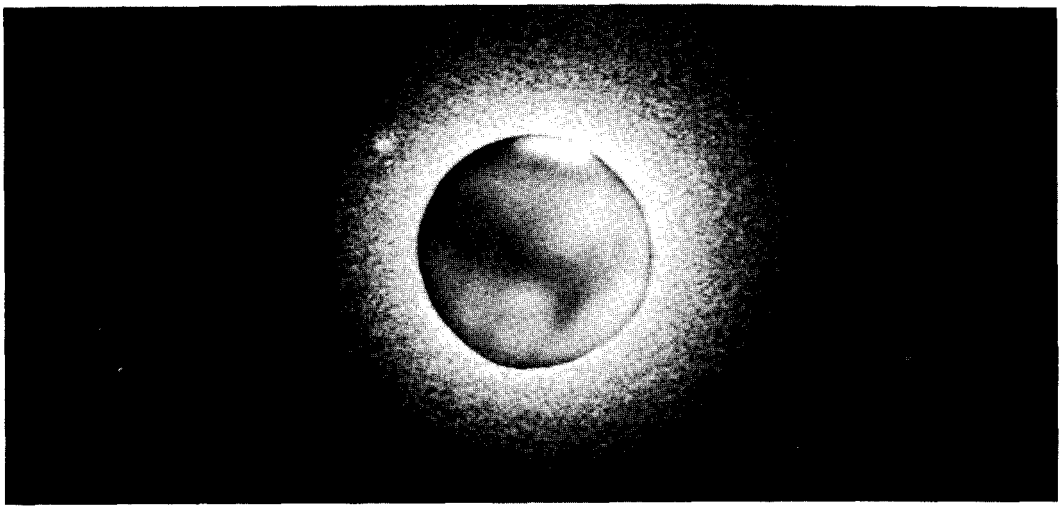


FIGURE 4.6. The satellites of Mars; the image of the planet was superimposed. (Lowell Observatory photograph.)

Mars; this is almost as close as is theoretically possible. According to the calculations made by the French astronomer E. A. Roche in 1849, a satellite cannot approach closer to the center of its parent planet, having the same density, than 2.44 times the radius of the planet. At shorter distances, tidal forces would overcome the gravitational adhesion and the satellite would disintegrate.

Like the Moon, Phobos revolves around its parent planet in the same direction as the planet rotates (fig. 4.7). The sidereal period of revolution, the time required to make a complete orbit, is 7 hours 39 minutes, less than a third the length of a Martian day. Phobos is the only known satellite in the solar system with a period of revolution that is shorter than the rotational period of its parent planet.

In this connection, Asaph Hall wrote:

> For several days the inner moon [Phobos] was a puzzle. It would appear on different sides of the planet in the same night, and at first I thought there were two or three inner moons, since it seemed . . . improbable that a satellite should revolve around its primary [i.e., parent planet] in less time than that in which the primary rotates. To decide this point I watched this moon throughout the nights of August 20 and 21, and saw that there was in fact but one inner moon, which made its revolution around the primary in less than one-third the time of the primary's rotation, a case unique in our solar system.

Because Mars rotates in the same direction as Phobos revolves in its orbit, the synodic period of the satellite, the time between two successive appearances at the same place in the Martian sky (p. 37), about 11 hours 6 minutes, is longer than the sidereal period.

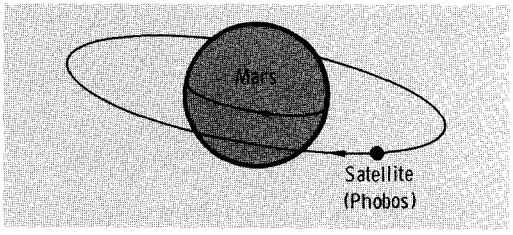


FIGURE 4.7. The motion of Phobos (south at top).

Thus, to an observer on Mars, the moon Phobos would appear to rise and set twice each day. The interval between two successive moonrises or moonsets is the synodic period of a little over 11 hours. Because the period of revolution of the satellite is less than the rotational period of the planet, the observer would see Phobos rise in the west and set in the east. Artificial Earth satellites with sidereal periods of less than 1 day behave in the same manner.

As a result of the proximity of Phobos to its parent planet and of the fact that its orbital plane is inclined at an angle of only 1.13 degrees to the equatorial plane of Mars, the satellite is always below the horizon at high latitudes in both northern and southern hemispheres. Thus, because of the curved surface of the planet, an observer on Mars at a latitude above about 70° N or S could never see the inner moon.

In 1945, the American astronomer B. P. Sharpless reported that an analysis of the motion of Phobos indicated the orbit of this satellite to be slowly shrinking; it appeared as if this moon were gradually spiraling inward toward Mars. Although G. M. Clemence called attention to the doubtful accuracy of the measurements, the matter attracted some interest, especially when the Soviet astrophysicist I. S. Shklovskii suggested in 1959 that Phobos was hollow and was presumably of artificial origin.

From a careful analysis of the observational data obtained at various oppositions from 1877 on, G. A. Wilkins of the Greenwich Observatory in England concluded in 1966 that there is strong evidence against an abnormality in the orbit of Phobos. Nevertheless, he felt that it could not be ruled out completely. Most of the difficulties in interpreting the data arise from errors in the observations, and a final decision must await developments in the theory of orbital motion.

The orbit of Deimos, with an eccentricity of 0.0028, is even more circular than that of Phobos. The orbital radius is some 23 500 kilometers (14 500 miles) and the outer moon is thus roughly 20 000 kilometers (12 000 miles) from the surface of Mars. The sidereal period of revolution of Deimos, also in the same direction as the rotation of the parent planet, is 30 hours 18 minutes, and the synodic period is about 5½ days.

Because Deimos revolves around Mars in a time period that is only a few hours longer than the planet's period of rotation, the outer moon would appear to a Martian observer to move slowly across the sky, rising in the east and setting some 2½ days later in the west. Deimos would then be on the other side of Mars and would not be seen for about another 3 days. The time between two successive risings or settings of Deimos is the synodic period of 5½ days.

The angle of inclination of the orbital plane of Deimos to the equatorial plane of Mars varies to some extent as a result of the disturbing action of the Sun. The angle is always small, however, ranging from 0.85 to 2.69 degrees. Because Deimos is farther from Mars than is Phobos, the former can be seen up to latitudes of about 83° N and S. At higher latitudes, near the poles, both moons are always below the local horizon and would not be visible.

Because of their proximity to the parent planet and short periods of revolution, Phobos and Deimos frequently pass through the shadow of Mars cast by the Sun. To an observer on the planet, the moons would then suffer eclipse, Phobos about 1400 times and Deimos roughly 130 times each Martian year. The satellites also frequently pass directly between the Sun and Mars, but they are much too small to produce an eclipse of the Sun, as the Moon does when it is on (or close to) the line joining the Sun and Earth. At these times, a Martian observer would see the

Planetary Characteristics of Mars

Mass	6.419×10^{23} kilograms	
Mass of Mars/mass of Earth	0.107	
Radius (equatorial)	3400 kilometers	2110 miles
Radius (polar)	3360 kilometers	2090 miles
Radius (average)	3380 kilometers	2100 miles
Radius of Mars/radius of Earth	0.531	
Flattening (optical)	0.011	
Flattening (dynamic)	0.00525	
Density	4 grams per cubic centimeter	
Gravitational acceleration (average)	375 centimeters per sec per sec	
Day (sidereal)	24 h 37 min 22.67 sec	
Day (solar)	24 h 39 min 35.23 sec	
Speed of rotation	0.26 kilometer (0.16 mile) per second	

moons as black spots moving rapidly across the face of the Sun. Such a phenomenon is more correctly described as a transit than as an eclipse.

Exploration of the Martian Satellites

Attention was drawn in 1966 by S. F. Singer in the United States to the potential interest of Phobos and Deimos as subjects for space exploration. These satellites were either produced at the same time as Mars or they are captured asteroids. In the former case, they would probably represent the initial material from which the terrestrial planets, Mercury, Venus, Earth, and Mars, developed. Because of the small size of the satellites, and the high rate of loss of heat by radiation, no important changes in physical or chemical nature are expected as a result of an increase in temperature. If, on the other hand, Phobos and Deimos were originally asteroids, then an opportunity would be provided for studying these members of the solar system.

At the present time, little more is known about the Martian satellites than the characteristics of their orbits. Approximate estimates have been made of their size, but their shape, mass, density, composition, surface characteristics, and rotation are completely unknown. A spacecraft orbiting Mars could get close enough to one or the other (or possibly both) of the satellites to determine the size, shape, and rate of rotation. Surface features could be observed by means of television cameras and the color and reflectivity might provide information concerning the composition. In the words of Singer, "a direct examination of the Martian satellites will shed important light on the origin and early history of the solar system."

FOR FURTHER INFORMATION

Ley, W.; von Braun, W.; and Bonestell, C.: The Exploration of Mars. The Viking Press, 1956.

Loomis, A. A.: Some Geologic Problems of Mars. Geol. Soc. Am. Bull., vol. 76, 1965, p. 1083.

Michaux, C. M.: Handbook of the Physical Properties of the Planet Mars. NASA SP–3030, 1967.

Öpik, E. J.: Atmosphere and Surface Properties of Mars and Venus. Progr. Astronaut. Sci., vol. 1, 1962, p. 261.

Pollack, J. B.; and Sagan, C.: An Analysis of Martian Photometry and Polarimetry. Smithsonian Astrophys. Obs. Spec. Rept. 258, 1967.

Vaucouleurs, G. de: Geometric and Photometric Parameters of the Terrestrial Planets. Icarus, vol. 3, 1964, p. 187.

Wells, R. A.: A Re-evaluation of W. H. Wright's Plates of the 1924 and 1926 Oppositions of Mars. Planetary and Space Sci., vol. 13, 1965, p. 361.

V

The Atmosphere of Mars

PLANETARY ATMOSPHERES AND THEIR ORIGIN

Introduction

A KNOWLEDGE OF THE CHARACTERISTIC properties of the Martian atmosphere is, of course, of scientific interest, both for its own sake and for the light it may throw on the origin of Earth's atmosphere. It has also, however, a practical significance. Within the foreseeable future, a vehicle containing scientific instruments, to be released from a spacecraft, will be landed on the surface of the planet. To slow down the lander, by the use of retrorockets or a parachute, or both, so that it does not crash on the surface, the composition, temperature, and pressure of the atmosphere of Mars, and their variation with altitude, should be known. Recent studies have shown that the atmospheric pressure at the surface of the planet is much lower than had been accepted for some 20 years. As a result, preliminary plans for the design of a landing vehicle have undergone extensive revision.

In the year 1784, William Herschel reported his opinion (p. 17) that Mars has a "considerable" atmosphere, although this view was challenged a few years later. There is no doubt that Mars has an atmosphere, but it is very much less dense than Earth's atmosphere. Moreover, its composition appears to be quite different. Until suitable instruments can be carried by spacecraft into the Martian atmosphere (ch. XIII), definite information concerning the identity of the gases present can be obtained only from spectroscopic studies of sunlight reflected from Mars as observed from Earth.

Identification of Gases by Absorption Spectra

Sunlight passes through the Martian atmosphere, first to the planet, and then after reflection by the surface and the atmosphere. In each passage the molecules and atoms present in the atmosphere of Mars absorb radiations of characteristic wavelength, and it is by the study of these absorption spectra on Earth that information concerning the composition of the Martian atmosphere has been obtained. Of course, if a particular species does not absorb any of the radiation from sunlight or the absorbed radiation is in such a region of wavelength that it does not reach the

instruments on Earth, there is no way of determining whether that species is present or not.

The spectroscopic study of Mars was initiated in 1862 by the English astronomer William Huggins, only 3 years after the invention of the method of spectral analysis. Shortly thereafter, similar work was done by others in the United States, in Germany, and in Italy, but in all cases the results were negative. The only absorption lines detected were those present in solar radiation, thus proving, at least, that the light from the planet is reflected sunlight.

Because of their great potential interest, spectroscopic investigations of Mars were continued throughout the latter part of the 19th century and into the 20th century. It was not until 1947, however, that the first definitive results were obtained. The difficulties in these studies arose mainly from the spectral lines normally present in sunlight, the absorption of radiation by gases in the terrestrial atmosphere, the weakness of the light from Mars that reaches instruments on Earth's surface, and inadequacies of the instrumentation. It is of interest that, as recently as the early 1960's, some absorption bands, which seemed to be of great significance when discovered in 1956, were shown to have originated in a rare constituent (HDO) of Earth's atmosphere rather than in Mars (p. 220).

Planetary Atmospheres

Before going on to describe what is known and what is surmised about the atmosphere of Mars, some of the general ideas concerning planetary atmospheres will be reviewed. By far the most abundant constituent element of the universe is hydrogen, and this is followed by helium. Next in order, although much less abundant than helium, are carbon, nitrogen, and oxygen.[1] It is widely accepted that the primitive atmospheres of the planets some 4.6 billion years ago consisted mainly of hydrogen and helium. It is further surmised that there were smaller amounts of gaseous compounds of hydrogen with carbon, nitrogen, and oxygen: methane (CH_4), ammonia (NH_3), and water vapor (H_2O). There may also have been small quantities of other gases, but those mentioned are thought to have been the most important. It may be noted that helium forms no known compounds, and so this gas would be present only in its elemental state.

It is highly probable that the atmospheres of the larger planets—Jupiter, Saturn, Uranus, and Neptune—consist chiefly of hydrogen and helium with some methane and ammonia. For example, the characteristic absorption lines (bands) of molecular hydrogen, methane, and ammonia have been detected in the spectrum of Jupiter, and there are reasons for believing that helium is also present.

The composition of Earth's atmosphere is, however, quite different from that of Jupiter (as is apparently the case, also of the smaller planets, Mercury, Venus, and Mars). The terrestrial atmosphere, for example, consists of 78.08 molecular (or volume) percent of nitrogen, 20.95 percent of oxygen, 0.93 percent of argon, with minor proportions of carbon dioxide (0.033 percent), water vapor (variable), and some other gases. Significant amounts of helium are present and there are traces of hydrogen, methane, and ammonia, but it is doubtful if these are of cosmological origin; that is to say, they probably do not originate from Earth's primitive atmosphere. Why then do the atmospheres of Earth and the smaller planets differ so markedly in composition from the atmospheres of Jupiter and the other large planets?

The situation can be accounted for by the

[1] The cosmic abundance of the chemically inert gas neon is probably similar to that of carbon, but, like helium, it is not important for the present discussion.

difference in the masses, and hence in the gravitational forces, of the large (Jovian- or Jupiter-like) and the small (terrestrial) planets. Because of the smaller gravitational attraction of the terrestrial planets, the lighter gases—gases of low molecular weight—such as hydrogen (molecular weight 2), helium (4), and neon (20), have escaped completely. The same is true of part of the methane (16), ammonia (17), and water vapor (18). In addition, ultraviolet radiation from the Sun can cause the decomposition of these three gases, and the resulting hydrogen will have escaped from the atmosphere in the course of time. The larger gravitational force on the Jovian planets and their greater distance from the Sun, on the other hand, have permitted the retention of the light gases of the primitive atmosphere.

The Origin of Earth's Atmosphere

From the foregoing, it may be concluded that the present atmosphere of Earth must be, at least in part, of secondary origin. It is thought at the present time that the nitrogen and some of the carbon dioxide in the atmosphere came from geochemical sources; the gases are formed by chemical action in Earth's outer layers and are then gradually released from the crust into the atmosphere. Some confirmation of this view is found in the observation that volcanic gases contain nitrogen, carbon dioxide, and water vapor, as well as small quantities of several other gases. Thus, the nitrogen gas, which constitutes almost four-fifths of the atmosphere, is regarded as having been released from Earth's crust by vulcanism, i.e., by volcanic or related action.

The next most important constituent, oxygen, is considered to have resulted from photosynthetic action by green plants. In the presence of sunlight (and moisture), such plants take up carbon dioxide gas and release oxygen. The overall process, which also involves incorporation of the carbon dioxide into complex compounds, like sugars and starches, is called photosynthesis, meaning synthesis by light (p. 214). Before the evolution of photosynthetic plants, Earth's atmosphere probably contained only very small proportions of oxygen produced by the decomposition of water vapor by ultraviolet light from the Sun. After green plants began to develop, perhaps a billion or so years ago, the proportion of oxygen in the air increased considerably, so that now it constitutes about one-fifth of the terrestrial atmosphere.

The third important constituent of the atmosphere is argon, present to the extent of almost 1 percent. This gas originates mainly from the radioactive decay of a form (isotope) of the element potassium having a mass number (atomic weight) of 40. All soils on Earth contain potassium compounds and the radioactive potassium-40 present decays slowly and is thereby partly converted into argon gas which enters the atmosphere.

The water vapor now in Earth's atmosphere is produced by evaporation from the bodies of liquid water which cover a large part of the globe. Some of this water may be of primitive cosmological origin. If there were low temperatures in the upper atmosphere, such as exist at present, they would act as a trap by freezing the water vapor and would thus inhibit its escape. Furthermore, in a moderate depth of liquid water, the ultraviolet radiation would be absorbed in the upper layers and the water below would be protected from decomposition. Much of the water now on Earth's surface, however, has probably resulted from the condensation of steam released from the interior by volcanic or related action. This steam may have been produced by the decomposition of hydrated, that is, water-containing, minerals by heat.

Most of the carbon dioxide in the present

atmosphere of Earth apparently came from the interior. An important factor in controlling the abundance of this gas is considered to be the chemical reactions which can occur between carbon dioxide and magnesium and calcium silicates in various minerals. These reactions, such as

$$MgSiO_3 + CO_2 = MgCO_3 + SiO_2,$$

lead to the formation of the respective carbonates plus silica. They occur fairly rapidly in liquid water, but only slowly in the presence of water vapor. In the complete absence of water in any form, the reactions of silicates with carbon dioxide take place extremely slowly, if at all. Other factors influencing the proportion of carbon dioxide in Earth's atmosphere are photosynthesis, combustion of coal, oil, and related fuels, and dissolution of the gas in the oceans.

THE MARTIAN ATMOSPHERE

Carbon Dioxide in the Martian Atmosphere

Spectroscopic studies of Mars, made over a period of more than 80 years, had produced no conclusive evidence concerning the nature of the atmospheric gases. All that could be said was that if certain gases were present, then their abundances must be less than various specified amounts. But there was no positive proof that these gases were present at all.

Although it was quite certain that Mars had an atmosphere, and approximate estimates had been made of its pressure, nothing was definitely known about its composition until near the end of 1947. On October 7 of that year, G. H. Kuiper, then at the McDonald Observatory, Texas, compared the infrared spectrum of Mars with that of the Moon obtained at the same time. The purpose of this comparison was to allow for spectral absorption in Earth's atmosphere. Since the Moon has no atmosphere, any difference in the spectra of Mars and the Moon could then be attributed to gases present in the Martian atmosphere.

The spectra obtained in this manner are reproduced in figure 5.1. The spectrum of Mars shows two small dips at wavelengths of 1.57 and 1.60 μ, with indications of others at 1.96, 2.01, and 2.06 μ, which are not present in the spectrum of the Moon.[2] These absorption wavelengths are characteristic of carbon dioxide, so Kuiper concluded that this gas is present in the Martian atmosphere. Confirmation was obtained by a more detailed observation of the spectrum made in February 1948 when Mars was closer to Earth. The three absorption bands in the wavelength region of 2 μ were more pronounced than in the original spectrum in figure 5.1. From the extent of the spectral absorption, Kuiper estimated the mass of carbon dioxide over unit area of the surface of Mars to be about twice the average over Earth's surface.

On Earth, atmospheric (barometric) pressures are commonly stated in terms of millimeters or centimeters, or inches, of mercury. A standard atmosphere at sea level on Earth, for example, is 760 millimeters of mercury. But 760 millimeters of mercury on Mars would not be the same pressure as a standard atmosphere on Earth because of the difference in the gravitational forces on the two planets. Consequently, other units are used to express pressures both on Earth and on the planets so that they can be compared with one another.

One such unit in common use is the bar, defined as a pressure of 1 million dynes per square centimeter (dyn/cm^2). (The dyne

[2] The symbol μ stands for 1 micron, which is one-millionth part of a meter. It is a convenient unit for expressing wavelengths in the infrared region of the spectrum.

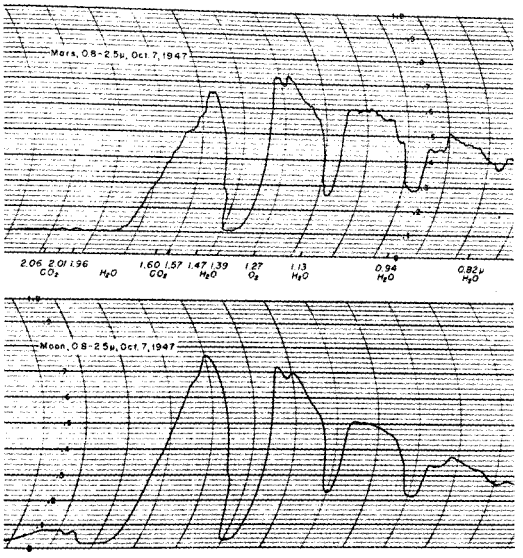


FIGURE 5.1. Infrared spectra of Mars and the Moon. (Spectra by G. P. Kuiper; courtesy University of Chicago Press.)

is a unit of force and pressure is force per unit area.) It happens that the standard atmosphere of 760 millimeters of mercury is equivalent to 1.013 million dyn/cm^2, or 1.013 bars. When making rough comparisons, 1 bar may be taken as approximately equivalent to the atmospheric pressure at Earth's surface. When it is required to indicate pressures of a smaller order of magnitude, it is convenient to employ the millibar, which is a one-thousandth part of a bar, as the unit. Thus, the standard atmospheric pressure on Earth is roughly 1000 (actually 1013) millibars.

Another way of stating the pressures or, rather, the abundances of particular atmospheric gases is in terms of their effective depth or reduced thickness. The reduced thickness of a particular gas is equal to its hypothetical thickness at a uniform density. This quantity of gas is taken to be that which would, at normal (terrestrial) atmospheric temperature and pressure, exert the same force on the surface as the actual gas in the planet's atmosphere. The unit commonly used is the meter-atmosphere, and for carbon dioxide (molecular weight 44) on the surface of Mars, the reduced thickness can be related to the pressure by

$$1 \text{ meter-atmosphere} = 0.074 \text{ millibar}$$

The abundance of carbon dioxide in the terrestrial atmosphere is equivalent to a reduced thickness of roughly 2 meter-atmospheres. Hence, according to Kuiper's estimate given above, the reduced thickness of the gas in the Martian atmosphere would be about 4 meter-atmospheres. A recalculation of the data, made by J. Grandjean and R. M. Goody in 1955, indicated that the quantity of carbon dioxide on Mars was more than six times greater than Kuiper had thought. This would make the reduced thickness about 27 meter-atmospheres, and the contribution of carbon dioxide to the pressure at the surface of Mars, that is, the partial pressure of this gas, would be roughly 2 millibars.

At the time these estimates were made, the total atmospheric (or barometric) pressure on Mars was considered to be in the vicinity of 80 millibars, and it appeared that carbon dioxide gas would constitute only a few percent of the atmosphere. The other major constituents would then presumably be gases that could not be identified because they did not produce detectable spectral absorption lines (or bands). Two such gases are nitrogen and argon which have absorption lines (or bands) in the short-wavelength (ultraviolet) region that cannot pass through Earth's atmosphere.

As mentioned earlier, nitrogen, presumably released from Earth's outer layers, is the chief constituent of the terrestrial atmosphere. Because it is relatively inert chemically and has a moderately high molecular weight (28), it

seemed reasonable to suppose that, in spite of the smaller gravitational attraction, nitrogen would also be found in the Martian atmosphere. Furthermore, argon, which is even more inert chemically and has a higher molecular weight (40) than nitrogen, might well be expected to be present, also. As on Earth, the argon would be produced by radioactive decay of potassium-40 in the outer layers of Mars.

In 1961 a typical estimate of the proportions of the major constituents of the Martian atmosphere was the following: 93 molecular percent of nitrogen, 5 to 6 percent of argon, and 1 to 2 percent of carbon dioxide. There would undoubtedly be some minor (trace) constituents, as in the terrestrial atmosphere. Among these, water vapor was considered to be a possibility, partly because of the existence of the polar caps. It was not until 1963 that the presence of water vapor in the Martian atmosphere was definitely established.

A dramatic change in the views concerning the composition of the atmosphere of Mars resulted from a study of the infrared spectrum made during the opposition in the early part of 1963. It appeared that the abundance of carbon dioxide was greater than had been accepted previously, about 55 meter-atmospheres, corresponding to a partial pressure of roughly 4.2 millibars at the surface. At the same time, the total atmospheric pressure on Mars was estimated to be in the vicinity of 25 millibars, rather than 80 millibars or so.

From these results the conclusion drawn was that the Martian atmosphere contains something like $(4.2/25) \times 100$, or about 17 percent of carbon dioxide. Subsequent spectral investigations have indicated that the reduced thickness of this gas on Mars may approach 100 meter-atmospheres; its partial pressure would then be approximately 7.5 millibars. At the same time, the estimates of the total pressure have been reduced somewhat, and values ranging from about 7 to around 20 millibars have been suggested.

The current view is that the atmosphere of Mars contains at least 50 percent of carbon dioxide. On the basis of the lower value of the atmospheric pressure given, some scientists have, in fact, suggested that the Martian atmosphere consists almost entirely of carbon dioxide. It may be significant, too, that a reconsideration of the computations made by Grandjean and Goody, based on Kuiper's approximate data of 1948, but using more recent values of the total atmospheric pressure, leads to the conclusion that carbon dioxide constitutes from 50 to 100 percent of the Martian atmosphere.

Water Vapor in the Martian Atmosphere

In addition to carbon dioxide, water vapor has been definitely identified in the atmosphere of Mars. This substance is of special interest because of its connection with the development of life on the planet. That water vapor is present has been assumed for many years. The brilliant white polar caps of Mars have long been thought to consist, at least in part, of small ice (or hoarfrost) crystals. Because one cap increases in size during the local autumn and winter in its hemisphere while the other is decreasing during the local spring and summer in the other hemisphere, it appears that there is a transfer of water vapor from one hemisphere to the other. If this is the case, there should always be some water vapor in the atmosphere of Mars.

During the early years of this century, observers claimed to have detected some of the characteristic infrared bands of water vapor in the Martian spectrum. Later work cast doubt on the validity of the results, and for some time the matter remained uncertain. Estimates of the amount of water vapor in

the atmosphere were made on the basis of various meteorological phenomena, such as the disappearance of morning haze, the change in size of the polar caps, etc., but there was no really positive evidence that there was any water in the atmosphere.

In 1963, however, L. D. Kaplan, G. Münch, and H. Spinrad, of the California Institute of Technology and the Jet Propulsion Laboratory, obtained definite spectroscopic proof that there is water vapor in the atmosphere of Mars. The positive detection was made possible by observation of the Doppler shift of the lines in the Martian spectrum. It is a familiar fact that the pitch (frequency) of a train whistle or automobile horn appears first to increase and then to decrease as the vehicle passes by a stationary observer. The change in frequency (wavelength), as a result of the relative motion of the source and the observer, applies to light as well as to sound.

The general phenomenon was first described by the Austrian physicist C. J. Doppler in 1842 and is called the Doppler effect. As a result of this effect, the wavelengths of all lines in the spectrum of Mars, as measured on Earth, are increased when the two planets are moving away from each other. At, or close to, eastern quadrature (fig. 3.9), Mars appears to be receding from Earth at its maximum speed of about 15 km/sec (24 mps). The conditions should then be optimum for observing the Doppler effect on the lines in the spectrum of Mars, and this could be used as an aid to identifying their source.

The spectrum of Mars, taken at the Mount Wilson Observatory on the night of April 12–13, 1963, when the planet was at eastern quadrature, after the opposition of February 4, 1963, showed the presence of a band in the wavelength vicinity of 8200 Å.[3] This band was caused by absorption by water vapor in Earth's atmosphere, but alongside each of the lines in the band was a weaker absorption line, shifted in wavelength by a mean value of 0.42 Å. The wavelength shift λ_D, caused by the Doppler effect, can be readily calculated from the expression

$$\lambda_D = \lambda \frac{v}{c}$$

where λ is the unshifted wavelength, v is the velocity with which Mars appears to be receding from Earth, and c is the velocity of light. In the present case, λ is 8200 Å, v is 15 km/sec, and c is 3.0×10^5 km/sec. Therefore, λ_D should be 0.41 Å.

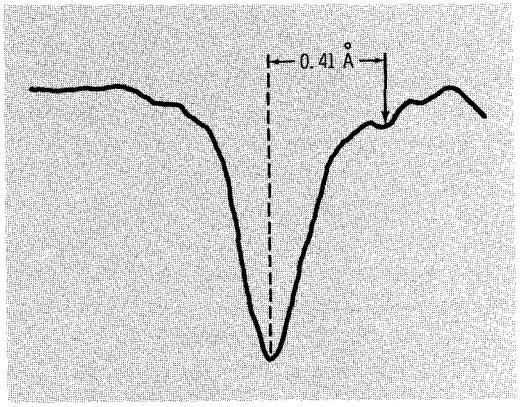


FIGURE 5.2. Doppler shift in the spectrum of water vapor near 8200 Å. (After L. D. Kaplan, G. Münch, and H. Spinrad.)

The shape of the spectral absorption curve, averaged over four lines of different wavelengths in the vicinity of 8200 Å, is seen in figure 5.2. The vertical line indicates the wavelength in the normal spectrum of water vapor, and which probably arises from absorption in Earth's atmosphere. The arrow

[3] The symbol Å represents the angstrom unit, equal to 10^{-8} centimeters, which is commonly used to express wavelengths of spectral lines.

shows where the Doppler-shifted line in the Martian spectrum should appear according to the calculation made above. The slight dip in the curve, located very close to the position of the arrow, is thus evidently caused by absorption by water vapor in the atmosphere of Mars.

It is the general practice to express the water vapor abundance on Mars in terms of the precipitable water. This is the hypothetical average (and uniform) depth of liquid water that would be formed over the whole planet if all the water vapor in the atmosphere were condensed into liquid water.

The quantity of water on Mars was calculated from the spectroscopic measurements (fig. 5.2) to be 1.4 ± 0.7 mg/cm^2. Because liquid water has a density of 1 g/cm^3, a mass of 1 mg/cm^2 corresponds to a depth of 10^{-3} centimeter of liquid water. The estimated precipitable water in the atmosphere of Mars is thus $(1.4\pm0.7)\times10^{-3}$ centimeter. Such small amounts are often stated in microns; as defined in the footnote on page 76, 1 micron is 10^{-6} meter, or 10^{-4} centimeter. Hence, the average amount of precipitable water on Mars was found to be 14 ± 7 microns.

For comparison, it may be noted that the amount of precipitable water is around 1000 microns over desert areas on Earth and it is up to 10 times as large where the humidity is high. As the average over the whole Earth is about 2000 microns, the Martian atmosphere is extremely dry from the terrestrial point of view.

The observations on the 8200-Å band described above were repeated and confirmed at the McDonald and Lick Observatories during the Martian apparition in the winter of 1964–65. Separate spectrograms were taken of the northern and southern hemispheres of the planet, and a definite connection was found between the estimated quantities of water vapor in the atmosphere and the changes in size of the polar caps. Detectable amounts of water vapor, approximately 15 microns of precipitable water, began to appear in the northern hemisphere in the late Martian spring when the dimensions of the north polar cap started to decrease rapidly. When the rate of shrinkage had leveled off, in late spring, water vapor was found to be distributed almost equally (10 microns) over both northern and southern hemispheres. At this time, water vapor was beginning to concentrate over the southern hemisphere, and by early summer in the northern hemisphere (early winter in the southern hemisphere) most of the water vapor (25 microns) was in the atmosphere over the southern hemisphere of Mars. These results suggest that water is transferred as vapor back and forth from one polar cap to the other according to the season.

The period during which reliable spectroscopic observations can be made for any given Martian apparition is very limited. Consequently, a series of studies of this kind at several opportunities will be required to determine how the water vapor in the atmosphere in the two hemispheres varies throughout the whole Martian year. It may be mentioned that on Earth there is relatively little transfer of water vapor between the Northern and Southern Hemispheres. For one thing, the changes in the polar caps of Earth are much less drastic than on Mars, and for another, there is a great deal more water available in each Hemisphere.

An alternative procedure for eliminating or reducing the effect of Earth's atmospheric water vapor is to make spectroscopic measurements at high altitudes. The radiations reaching the instruments have then passed through only a small part of the atmosphere, for which allowance can be made. Observations of the infrared spectrum of Mars, at a wavelength of 1.4μ, were carried out in January 1963 from a balloon at an altitude of 14 kilometers

(45 000 feet) over France and also from the Jungfrau Scientific Station, Switzerland, at an elevation of 3.5 kilometers (11 200 feet). The results, according to A. Dollfus, indicated that the precipitable water on Mars was about 150 microns. This value, however, is thought to be too high.

In March 1963, the unmanned balloon Stratoscope II, carrying a 36-inch reflecting telescope and an infrared spectrometer, was launched from Palestine, Tex., and reached an altitude of some 80 000 feet. The absorption in the spectrum in the wavelength region of 2.7μ was in best agreement with 10 microns of precipitable water in the atmosphere of Mars, but uncertainties in the reduction of the data did not permit a more precise conclusion than that the quantity of water was less than 40 microns.

Other Gases in the Martian Atmosphere

It is expected that there are small quantities of carbon monoxide and of oxygen (atoms and molecules), particularly in the upper layers of the Martian atmosphere, produced by the decomposition of carbon dioxide, viz

$$CO_2 \rightarrow CO + O \qquad \text{and} \qquad O + O \rightarrow O_2$$

by the ultraviolet radiation of wavelength shorter than 1600 Å in sunlight. As in Earth's atmosphere, some of the oxygen atoms (O) and molecules (O_2) would probably combine to form ozone (O_3).

During 1968, an unpublished (oral) report from the Jet Propulsion Laboratory implied that spectroscopic evidence had been obtained for the presence of carbon monoxide in the atmosphere of Mars. Although this result has not yet been confirmed, there is little reason to doubt its accuracy. If there is indeed some carbon monoxide in the atmosphere of Mars, then traces of oxygen and ozone are probably also present. So far, however, they have not been identified.

Spectral observations of Mars are made at all possible opportunities with the object of detecting constituents of the atmosphere, in addition to those mentioned above. So far, negative results have been reported for methane (CH_4), ammonia (NH_3), nitrogen dioxide (NO_2), hydrogen sulfide (H_2S), formaldehyde (HCHO), and acetaldehyde (CH_3CHO).

A preliminary report, published in 1966, indicated that there might be traces of certain hydrocarbons, that is, compounds of hydrogen and carbon, known as substituted methanes, in the Martian atmosphere. If this had been correct, it would have had considerable significance in connection with the problem of the occurrence of life on the planet, but a careful review of the spectroscopic data showed that the results were probably caused by instrumental factors.

Among the gases which have been sought but not yet detected in the Martian atmosphere, special interest attaches to nitrogen dioxide. In 1963, C. C. Kiess and his associates in the United States claimed to have observed some of the characteristic lines of this molecule in the spectrum of Mars. On the assumption that significant amounts of nitrogen dioxide are present, they had previously (1960) developed novel interpretations of several Martian phenomena, including the nature of the polar caps and the yellow clouds and the cause of the so-called blue haze.

Because of its importance in these respects, several scientists have searched for the lines of nitrogen dioxide in the spectrum of Mars, but without success. It has been estimated, in fact, that the upper limit of the abundance of nitrogen dioxide in the Martian atmosphere is about 0.01 millimeter-atmosphere. This is much less than that normally found over the city of Los Angeles. Because

nitrogen dioxide is decomposed by the ultraviolet radiation in sunlight, it is not surprising that the amount, if any, in the atmosphere of Mars is very small.

Composition of the Martian Atmosphere

In reviewing what is known about the composition of the atmosphere of Mars, all that can be said is that carbon dioxide is probably the major constituent. There are very small quantities of water vapor and possibly also some carbon monoxide. If the latter is present, then traces of oxygen and ozone are to be expected, but they have not yet been detected.

If the atmosphere of Mars was produced by the liberation of gases from the outer layers (crust) of the planet, in a manner similar to that which apparently occurred on Earth, some nitrogen should be present. Estimates of the composition of the gases exhaled by Earth's crust indicate that they contain roughly 5 to 10 molecules of carbon dioxide to one of nitrogen. On Earth, most of the carbon dioxide has been removed by processes requiring the presence of liquid water. It is doubtful, in view of the small amount of water now available, that any of these processes could have taken place to a substantial extent on Mars.

Furthermore, because molecular nitrogen is a relatively inert gas with a moderately high molecular weight, the loss of this gas, by chemical reaction or by escape into space, should be small. It would seem, therefore, that the ratio of carbon dioxide to nitrogen in the present atmosphere of Mars should not be greatly different from that of the gases released from the crust. If this is the same as on Earth, then the Martian atmosphere may have some 10 to 20 molecular percent of nitrogen. From the available data on the abundance of carbon dioxide and the total atmospheric pressure on Mars, it would appear that this is possible but not certain.

It has long been assumed that the Martian atmosphere also contains some argon, perhaps a little over 1 percent of the quantity of nitrogen, as in Earth's atmosphere. Because there are undoubtedly potassium compounds near the surface of Mars, there will probably be some argon in the atmosphere produced by radioactive decay of potassium-40. On Earth, the differentiation processes, described in chapter IV, which have taken place in the interior have resulted in a concentration of potassium near the surface. If there has been no differentiation in the interior of Mars, then the proportion of potassium in the crust may be lower than on Earth, and then so also would be the abundance of argon in the atmosphere.

The foregoing discussion has been based on the assumption that the present Martian atmosphere has originated entirely from the interior of the planet. Some scientists, however, have suggested the possibility that part, at least, may have come from external sources, such as impacting meteorites and, especially, comets. According to one widely held theory, comets contain substantial quantities of solidified carbon dioxide, methane, ammonia, and water. Upon impact with a planet, these substances would vaporize and so contribute to the atmosphere. Because the mass of a large comet approaches that of the Martian atmosphere, which is about 3×10^{16} kilograms, the impact of such a comet could have a significant influence on the atmospheric composition.

Possibility of Permafrost on Mars

The small amount of water vapor in the Martian atmosphere presents an interesting problem. In terrestrial volcanic gases, water vapor is by far the most common constituent,

and the condensation of this vapor has led to the formation of the oceans. The interior of Mars probably contains chemically bound water in the form of hydrates and there is some evidence that such hydrated compounds are present on the surface of the planet. For example, as will be seen in the next chapter, the reddish-yellow color of Mars is generally ascribed to a hydrated oxide of (ferric) iron. Furthermore, according to a report published in 1967 by W. M. Sinton of the Lowell Observatory, Flagstaff, Ariz., a strong emission band at a wavelength of around 3.1μ in the Martian spectrum, from both light and dark areas, is produced by hydrated minerals.

As a result of heating in the interior of the planet, the water present in these hydrates should be liberated as vapor from the crust of Mars, just as it is on Earth. If the carbon dioxide and nitrogen in the Martian atmosphere have also been released from the interior, then they should have been accompanied by substantial amounts of water vapor. What, then, has happened to this water? Some has probably been decomposed into hydrogen and oxygen by ultraviolet radiation from sunlight. The hydrogen, being a very light gas, has escaped. The very small, so far undetected, quantity of oxygen in the Martian atmosphere suggests either that there has been very little decomposition of water, or more probably, that the oxygen has reacted chemically with materials on the surface. The apparent presence of highly oxidized (ferric) iron compounds (ch. VI) indicates that this may have occurred.

The intriguing suggestion has been made that water vapor from the interior of Mars is retained just beneath the surface in the form of permafrost. In regions of high latitude on Earth, in the vicinity of the poles, the temperature of the ground below a shallow depth is always less than 0° C. Water is then permanently frozen into the soil and becomes an essentially integral part of it. This is the permafrost which is estimated to occur under about a fourth of Earth's land areas. In some places, such as northern Alaska and Siberia, it has a depth of almost 1000 feet.

There are reasons for believing that, on Mars, the temperature a very short distance under the surface is always well below 0° C over almost the whole of the planet (ch. VI). The conditions are thus everywhere suitable for the trapping as permafrost of any water vapor released from the crust. In these circumstances, and because there are no bodies of water on Mars, the abundance of water vapor in the atmosphere would inevitably be low. Nevertheless, the total quantity of water just below the ground might be considerable. Some laboratory evidence for the formation of permafrost under Martian conditions is described in chapter X. As will be seen there, the matter has important implications in connection with the possibility of the existence of life on the planet.

The Atmospheric Pressure on Mars: Early Estimates

Prior to 1963, two main methods were used to estimate the (barometric) pressure of the atmosphere at the surface of Mars. One of these, called the photometric (light-measuring) method, depends on observations of the variation of the brightness of the planet at different wavelengths and phase angles (p. 43). The apparent brightness arises from the scattering, or random reflection (as distinguished from specular or mirrorlike reflection), of sunlight both by the surface material and by the atmosphere (fig. 5.3). From the observations mentioned above, the relative contributions of these two sources can be calculated. The amount of scattering from the atmosphere can thus be estimated. Then, by

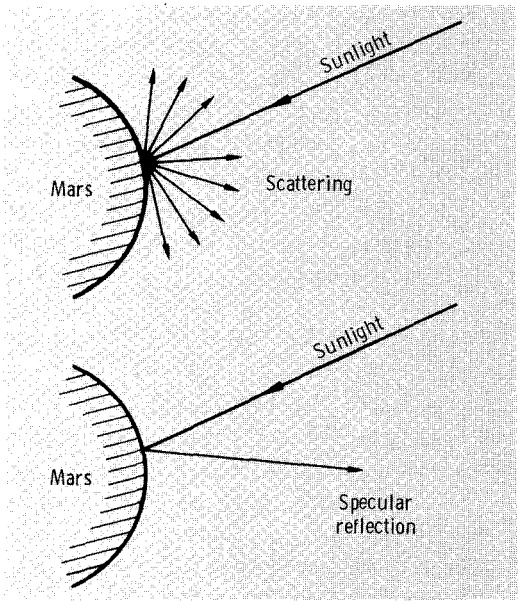


FIGURE 5.3. Scattering and specular reflection.

the use of established formulas relating the pressure of a gas to the brightness caused by scattering, the pressure of the Martian atmosphere can be evaluated.

One of the first, if not the first, attempt to use light-scattering data to estimate the barometric pressure on Mars was made in 1908 by Percival Lowell. He took the visual albedo (p. 65) of Mars to be 0.27, and from this he somewhat arbitrarily subtracted 0.10 for the surface, leaving 0.17 for the albedo of the atmosphere. Then, he estimated the albedo of Earth's atmosphere to be 0.75. The albedo values used by Lowell are now known to be too high, but his approach to the problem is nevertheless of interest.

The amount of scattering of light by a gas is roughly proportional to its mass per unit area. Because the albedo is a measure of the scattering, Lowell estimated that the total mass of the atmosphere per unit area of Mars is $0.17/0.75=0.23$ of the mass per unit area of the terrestrial atmosphere. The force of gravity on Mars is about 0.38 times that on Earth, so the atmospheric pressure on Mars should be $0.23\times0.38=0.087$ of Earth's barometric pressure. As seen earlier, the Earth's barometric pressure is about 1 bar, which means the pressure on Mars should be roughly 0.087 bar, or 87 millibars.

In 1926, D. H. Menzel in the United States used a more refined treatment of the albedo of Mars at different wavelengths to show that the ratio of the Mars/Earth atmospheric masses per unit area is about 0.18. The barometric pressure on Mars would then be approximately 68 millibars.

An important advance in the use of light-scattering measurements to determine the Martian atmospheric pressure was reported in 1934 by the Russian astronomers N. Barabashov and B. Semejkin. They used red, blue, and yellow filters to obtain photographic data on the scattering of light from bright areas on Mars at three different wavelengths. From the results the atmospheric pressure was estimated to be some 50 millibars. Later work by Barabashov and other Russian scientists, however, gave significantly higher pressures, over 100 millibars, using the same technique.

From visual estimates of the brightness of different surface areas, made in France during the Martian apparition of 1939, G. de Vaucouleurs developed simple procedures for determining the relative brightness of the surface and atmosphere of the planet. In this way, the atmospheric pressure was calculated to lie between about 80 and 90 millibars.

The second of the older techniques for determining the surface pressure on Mars is called the polarimetric method. It is considered to be capable of yielding more accurate results than the methods depending on light scattering. To understand its essential principle, it is necessary to say something about the polarization of light. Incidentally, polarization measurements have been frequently used

in studies of the surface and atmosphere of Mars. Relevance of this technique to surface studies is discussed in chapter VI.

Light (and similar electromagnetic radiations) can be regarded as consisting of electric and magnetic waves combined in such a manner that the electric and magnetic field directions are always at right angles. For the present purpose, it is sufficient to consider only the electric wave. Suppose that a plane is placed at right angles to the direction of propagation of the light. The electric wave will then intersect this plane at a line, as indicated by the arrow in figure 5.4. This intersection, which has both magnitude and direction, is called the electric vector of the wave.

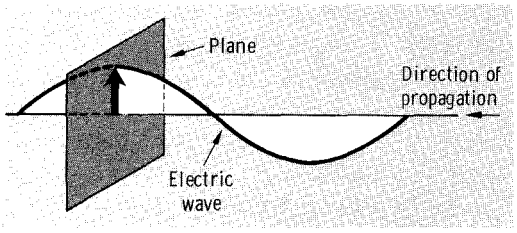


FIGURE 5.4. Electromagnetic wave and electric vector.

In unpolarized light, the electric vectors of successive waves are completely random and point in all directions, as shown in figure 5.5. In polarized light, however, there is a certain degree of order in the orientations of the electric vectors. Specifically, in plane polarized light, the electric vectors always lie in the same plane. By the use of suitable materials, such as tourmaline or calcite crystals or Polaroid sheet, the light which is polarized in a particular plane can be distinguished from that polarized in another plane. For the present purpose, it is important to note that scattered light is always polarized to some extent.

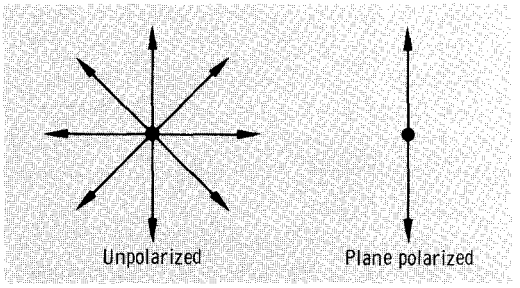


FIGURE 5.5. Directions of electric vector in unpolarized and plane polarized light.

Like the brightness of the planet, the polarization of the reflected scattered light from Mars is due in part to light scattered by the surface and in part to scattering by the atmosphere. The total polarization is expressed numerically by a quantity P which is defined by

$$P=\frac{I_1-I_2}{I_1+I_2}$$

I_1 is the intensity of the light polarized in a plane at right angles to the one containing the Sun (light source), Mars (scatterer), and Earth (observer); I_2 is the intensity polarized in a plane that is parallel to the Sun-Mars-Earth plane.

By utilizing polarization measurements on various powders, which simulate the Martian surface, the contribution of the atmosphere was calculated from the observed polarization of the light from Mars. The atmospheric polarization can then be related to the barometric pressure. The first attempt to use polarization measurements to evaluate the atmospheric pressure on Mars was made in 1929 by the French astronomer Bernard Lyot, a pioneer in the study of the Moon and planets by polarimetric methods. On the basis of the assumption that the Martian atmosphere consists mainly of nitrogen, he determined the

surface pressure to be not greater than 25 millibars.

Further development of the polarimetric method, utilizing measurements made at several apparitions, led Lyot's associate Adouin Dollfus to conclude in 1950, and again in 1957, that the barometric pressure on the surface of Mars was in the vicinity of 85 to 90 millibars. This result appeared to confirm the values derived from the best photometric measurements, and in a critical review, published in 1954, de Vaucouleurs gave 85 ± 4 millibars as "the most probable value of the atmospheric pressure at ground level on Mars."

Recent Determination of the Atmospheric Pressure

For several years, the surface pressure on Mars was accepted as 85 millibars and calculations of the best way of landing an instrument capsule on the planet were based on this atmospheric pressure. In 1963, observations on the infrared spectrum of carbon dioxide on Mars, to be described shortly, indicated that the value of the pressure given above was too high, perhaps by a factor of 4 or more. A reexamination of the older data, with an improved allowance for the polarization of the light scattered by the surface material, gave an atmospheric pressure of about 50 millibars. Then, on the basis of polarization measurements made in France by A. Dollfus and J. Focas during the 1965 opposition period, J. B. Pollack in the United States has estimated the surface pressure to be 15 millibars if the atmosphere consists entirely of carbon dioxide and 19 millbars if it is half nitrogen. If, as some have suggested, allowance should be made for light scattered by very fine solid particles suspended in the Martian atmosphere, then the pressure would be lower still. There are, however, important differences of opinion on this point.

It is of interest that a new approach to the reflectivity method has also led to results for the Martian atmospheric pressure much lower than those given earlier. Both S. Musman (1964) and D. Evans (1965), working independently in the United States, have made use of reflectivity measurements in the ultraviolet region of the spectrum where the light scattered from the surface itself is essentially zero. The reflectivity can then be attributed almost entirely to the atmosphere and from it the surface pressure can be calculated.

By utilizing the Martian albedo at a wavelength of 3300 Å, as determined by G. de Vaucouleurs, Musman estimated the pressure to be 19 millibars for an atmosphere consisting entirely of carbon dioxide and 27 millibars for one of nitrogen. Evans, on the other hand, based his calculations on measurements of reflectivity in the ultraviolet spectrum of Mars, in the wavelength range of 2400 to 3500 Å, made from a rocket above Earth's tangible atmosphere. The results could be correlated with surface pressure values ranging from 5 to 20 millibars, for possible atmospheres consisting of various mixtures of carbon dioxide, nitrogen, and argon. The most probable pressure value was given as 10 millibars.

A new era in the study of the Martian atmospheric pressure was initiated in 1963 when H. Spinrad, working at the Mount Wilson Observatory, obtained a high-resolution spectrum of Mars in the near-infrared region; that is, at wavelengths slightly longer than visible light. Spinrad was actually looking for the spectrum of water vapor, which was observed, but in addition he noted the presence of weak bands caused by carbon dioxide. The lines in this band, located at wavelengths around 8700 Å, are very faint and their effective width is almost independent of the atmospheric pressure. From an analysis of the spectrum, reported early in 1964 by L. D.

Kaplan and G. Münch in conjunction with H. Spinrad, the abundance of carbon dioxide in the Martian atmosphere was determined to be equivalent to about 55 meter-atmospheres.

In contrast to the carbon dioxide band at 8700 Å, there is a much stronger infrared band at a wavelength of about 2 μ (20 000 Å), which has been well studied under laboratory conditions. In the latter band, the line widths are known to depend both on the abundance of the carbon dioxide and on the total pressure of the atmosphere. With the abundance derived above from the 8700-Å band, which is essentially independent of pressure, the atmospheric pressure at the surface of Mars was calculated to be roughly 25 millibars.

Confirmation for what appeared at the time to be a surprisingly low value of the Martian barometric pressure soon came from studies of the carbon dioxide infrared spectrum of Mars made by T. C. Owen and G. P. Kuiper at the Lunar and Planetary Observatory, University of Arizona, and by V. I. Moroz in the U.S.S.R. The former quoted a mean of 17 millibars for the pressure, whereas the latter gave about 15 millibars, although with a considerable probable error. Subsequently, other observers, using spectroscopic measurements, have reported atmospheric pressures on the surface of Mars in the general range of 10 to 20 millibars. As seen above, this is close to the range of values derived from the newer polarimetric and reflectivity calculations.

THE MARINER IV OCCULTATION EXPERIMENT

General Description of the Experiment

At the time when the earliest spectroscopic data for the Martian atmospheric pressure were beginning to appear, preparations were in hand for launching a Mariner spacecraft to Mars (ch. XIII). The trajectory of Mariner IV, which passed close to Mars on July 14, 1965, was deliberately chosen so that information could be obtained about the atmosphere of the planet. The spacecraft carried a transmitter which generated radio waves at a constant frequency (2300 million cps). These waves were normally utilized for tracking purposes and for telemetering information from the spacecraft instruments to receivers on Earth.

As Mariner IV passed close to Mars, the radio signal was transmitted first through the upper levels (ionosphere) of the atmosphere of Mars on its way to Earth (fig. 5.6). In due course, the signal passed successively through the lower layers of the Martian atmosphere, until the spacecraft went behind the planet, as viewed from Earth, and the signal was cut off. Mariner IV was then said to be occulted by Mars.

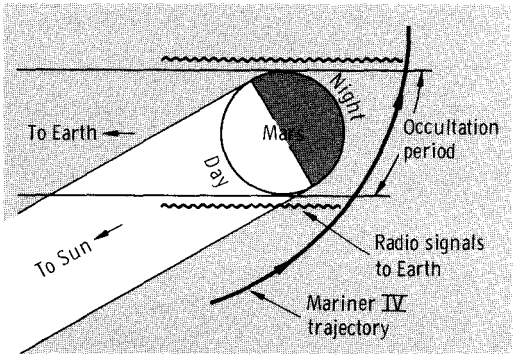


FIGURE 5.6. Occultation of Mariner IV by Mars.

The occultation lasted for almost 54 minutes, after which the signal from Mariner IV was again received on Earth. At first the signal passed through those layers of the Martian atmosphere closest to the surface. The later signals were transmitted through the

upper layers on their way to Earth. Before occultation, Mariner IV was on the daylight side of Mars, and the radio signals passed through the sunlit atmosphere of the planet. After occultation, the part of the atmosphere through which the transmission occurred was on the night side of Mars (fig. 5.6).

The radio signal from Mariner IV in its normal transmission to Earth—that is, when it was not passing through the Martian atmosphere—was affected by two factors. These are the motion of the spacecraft with respect to Earth and passage of the radio wave through Earth's atmosphere to the receiver. A radio signal passing close to Mars, however, is changed considerably because the path of the radio waves is bent (or refracted) by the Martian atmosphere (fig. 5.7). Furthermore, the extent of the bending (refractive index) of the path of the radio signal varies with the altitude at which it passes through the atmosphere of Mars on its way to Earth.

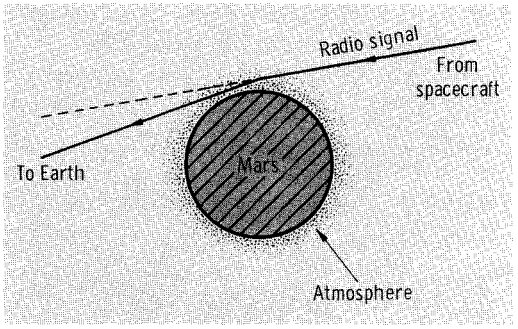


FIGURE 5.7. Refraction of radio signal by the atmosphere of Mars.

Results of the Occultation Experiment

From very precise measurements, made by receivers on Earth, on the signals transmitted by Mariner IV, both well before and after occultation by Mars, the characteristic signals expected in the interval were calculated. By subtracting these ideal theoretical values from the observations, the effect of the Martian atmosphere on the radio waves could be determined. The results, given as a change in the equivalent, or phase, path, expressed in cycles or wavelengths as a function of time, are reproduced in figure 5.8.

It is seen that at about 02:29:20 (Greenwich mean time), the electrically charged ionosphere of Mars is just beginning to affect the radio transmission. At this time, the signal was passing through a level in the atmosphere estimated to be at an altitude of about 250 kilometers (150 miles). Subsequently, this negative effect increased and attained its maximum magnitude indicated by the minimum in the curve at 02:30:12. The altitude of the level then being probed by the signal was roughly 120 kilometers (74 miles) from the surface. There is a small bend in the curve at 02:30:22, which is thought to be caused by passage of the radio signal through a lower level, just below 100 kilometers (62 miles) altitude, in the ionosphere.

The effect of the neutral, lower atmosphere on the transmitted signal commenced around 02:30:50 at an altitude of roughly 50

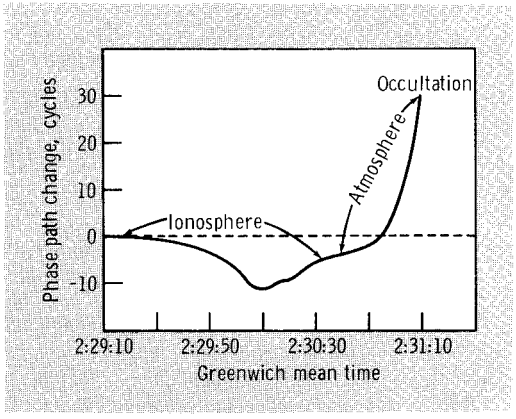


FIGURE 5.8. Phase path changes in radio signal from Mariner IV just prior to occultation.

kilometers (31 miles). From this point on, the phase path change curve in figure 5.8 rises rapidly, as the signal penetrates deeper and deeper into the increasingly dense atmosphere of Mars. Finally, shortly after 02:31:11, the signal from Mariner IV ceased suddenly because of the onset of occultation of the spacecraft by Mars.

After emergence from occultation, the signal phase difference changed with time in a reverse manner to that in figure 5.8. In other words, it started at the right and proceeded to the left. The atmospheric portion of the curve was very similar to that before occultation, but the minimum attributed to the ionosphere was absent. This result was not altogether surprising because the Martian ionosphere might well disappear at night.

The shape of the atmospheric part of the curve in figure 5.8 can be correlated with the radio-wave refractive index close to the Martian surface and to the scale height of the atmosphere. The scale height is a measure of how the refractive index decreases with the altitude, and it can be used to calculate the average gas temperature. From the refractive index, the number of gas molecules per unit volume (or molecular number density) can be estimated, provided the composition of the atmosphere is known. For calculation purposes, certain reasonable compositions are specified and the number density of the molecules is computed for each case. To determine the atmospheric pressure, the temperature is required and this can be derived from the scale height.

It can be shown that the scale height is numerically equal to RT/Mg, where R is the universal ideal gas constant (8.31×10^{7} ergs per mole per degree), T is the temperature in degrees Kelvin (°K=°C+273), M is the average molecular weight of the atmospheric gases, and g is the acceleration of gravity at the Martian surface (about 375 cm/sec/sec). If the atmosphere is assumed to consist entirely of carbon dioxide, M is 44. From the Mariner IV measurements, the scale height was estimated to be 8 to 10 kilometers; that is, 8×10^{5} to 10^{6} centimeters. Upon setting RT/Mg equal to these values, and using the data given above, the corresponding temperatures of the Martian atmosphere are found to be about 160° K and 200° K, respectively. The result is written as 180°±20° K for an assumed atmosphere of carbon dioxide.

The pressure at the surface is equal to kTC, where k is the Boltzmann gas constant per single molecule (1.38×10^{-16} erg per degree), T is the calculated temperature, and C is the molecular number density in the atmosphere at the surface. As seen above, the value of C is estimated from the refractive index for an assumed atmospheric composition. If k is in ergs per degree Kelvin, T in °K, and C in molecules per cubic centimeter, the pressures will be in dynes per square centimeter. Upon multiplication by 1000, the values will be in millibars.

The results obtained in the foregoing manner from the Mariner IV data are summarized in the table on the next page for two atmospheric compositions: 100 percent carbon dioxide and 80 percent carbon dioxide plus 20 percent nitrogen. The pressures would be somewhat higher if there were a larger proportion of nitrogen. The molecular weight of argon (40) is not very different from that of carbon dioxide (44). For this reason, the presence of significant quantities of this gas in the Martian atmosphere would have only a minor effect on the calculated pressures.

It will be apparent from the data in the table that there was a marked difference in the results before and after occultation. A possible explanation is that the surface of Mars grazed by the signal immediately upon emersion from occultation may have been at a

	Before occultation	*After occultation*
Scale height (kilometers)	9±1	12±1
Surface number density (10^{17} molecules/cm^3):		
100% CO_2	1.9±0.1	2.25±0.15
80% CO_2+20% N_2	2.1±0.1	2.45±0.15
Surface temperature (°K):		
100% CO_2	180±20°	240±20°
80% CO_2+20% N_2	170±20°	220±20°
Surface pressure (millibars):		
100% CO_2	4.9±0.8	8.4±1.3
80% CO_2+20% N_2	5.2±0.9	8.8±1.3

lower altitude, and hence had a higher barometric pressure, than the surface just before occultation. The higher atmospheric temperature in the former case would then be in general agreement with expectation for the lower altitude of the surface.

In addition to giving the variation of the radio-wave refractive index with altitude, the scale height is also equal to the vertical distance, at constant temperature, in which the number density of gas molecules changes by a factor of e (that is, 2.72), the base of natural logarithms. By assuming a scale height of about 10 kilometers, the two atmospheric pressure values of 5 and 8.5 millibars before and after occultation, respectively, would correspond to a difference in elevation of some 5 kilometers (16 400 feet).

The occultation method for determining atmospheric pressure is inevitably biased in favor of the more elevated regions in the area where the radio wave grazes the surface of the planet just before and after occultation. So the pressure of the 8.5 millibars does not necessarily represent the maximum barometric pressure at the lowest levels on Mars. In fact, on the basis of certain radar reflection observations described in chapter VI, C. Sagan and J. B. Pollack of the Smithsonian Institution Astrophysical Observatory estimated in 1966 that at the very lowest levels, corresponding roughly to the bottoms of deep oceans on Earth, the atmospheric pressures might be as high as 20 millibars. As already seen, values of 10 to 20 millibars have been derived from reflectivity, polarimetric, and spectroscopic studies.

Another point in connection with the Mariner IV data is worthy of mention. Just before occultation, the radio beam grazed a bright region of Mars at about latitude 50° S and longitude 183° W, between Electris and Mare Chronium. After occultation, it appears that a dark area was grazed near 60° N and 34° W, at the north of Mare Acidalium. If the explanation of the observed pressure differences suggested above is correct, the particular dark area on Mars is at a lower altitude than the bright area. C. Sagan and J. B. Pollack have indicated, however, that within the limits of possible error, the area grazed by the radio waves after occultation may have been a bright one adjacent to dark areas. This matter is of some importance, as will be seen in the next chapter.

It may be noted, in any event, that barometric pressures do not indicate altitudes above the ideal physical surface of a planet (or distance from its center), but rather above a hypothetical equipotential surface of gravity. This is defined as a surface on which no work is required to move a mass from one point to

another. The equipotential surface of Earth is very complex, with many undulations which are determined by the mass distribution below the surface of the planet. It is not at all unlikely that the equipotential surface of Mars is also complicated, and so it would be unsafe to draw conclusions concerning differences in altitude from atmospheric pressures measured at considerable distances apart. It is evident that much more needs to be known about Mars before the available data can be interpreted completely.

Variation of Atmospheric Pressure With Altitude

Before the Mariner IV occultation experiment, the only way in which the scale height of the Martian atmosphere could be determined was by calculating the value of RT/Mg, using an estimated average molecular weight and temperature. On the basis of the assumption, accepted before 1963, that the atmosphere consisted mainly of nitrogen molecules (molecular weight 28) at a temperature of 180° K, the scale height was calculated to be about 14.5 kilometers. This may be compared with a scale height of approximately 7 kilometers in Earth's atmosphere.

The smaller the scale height, the more rapidly does the molecular-number density in the atmosphere decrease with increasing altitude. Hence, the atmospheric number density on Earth should decrease more rapidly than on Mars. For several years it was widely accepted that, above an altitude of some 50 kilometers (31 miles), the number density in the Martian atmosphere was actually greater than at the same altitudes above Earth, although at the surface it is much larger on Earth than on Mars.

The question of the atmospheric density at high altitudes over Mars is of more than a purely scientific interest. In order to design an orbiting spacecraft and to determine its orbit for a specific lifetime, a knowledge of the atmospheric density is a fundamental requirement. The larger the mass density, the number density multiplied by the molecular weight, the greater the resistance to motion of an orbiting spacecraft, and hence the shorter will be its lifetime.

It now appears that the scale height of the Martian atmosphere at high altitudes is only about 9 kilometers, and this is not greatly different from the scale height of the terrestrial atmosphere. It is probable, therefore, that the atmospheric mass density on Mars, at all altitudes of interest, is less than on Earth. The estimated mass densities, expressed in grams per cubic centimeter, for altitudes up to 140 kilometers (87 miles) are shown in figure 5.9.

Because the mass densities are less than those based on the older data, it should be possible to place a spacecraft in a lower orbit around Mars than had previously been considered possible. With a lower orbit, better

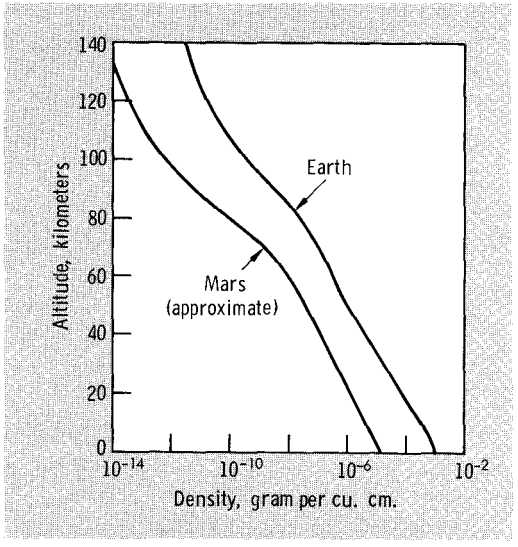


FIGURE 5.9. Variation of atmospheric densities of Mars (estimated) and Earth with altitude.

information should be obtainable about the planet. Although the low atmospheric pressure (and mass density) near the Martian surface will be a drawback for the landing of an instrumented capsule on the surface, the fact that the mass density is correspondingly low at high altitudes would be of advantage to a Martian orbiter.

The Ionosphere of Mars

In Earth's atmosphere above an altitude of about 60 kilometers (35 miles), there is a region known as the ionosphere, where there are significant numbers of atoms or molecules with a positive electrical charge—called positive ions—and electrons, with negative charges. In the daytime, four more-or-less distinct regions, at different altitudes, can be identified in the ionosphere, and these are designated by the symbols D, E, F_1, and F_2. The electron and ion densities in these regions and the dependence on altitude vary with the season and with the activity of the Sun, as well as with the time of day. For comparison with the results for Mars obtained from Mariner IV, the curves in figure 5.10 show the day and night electron-number densities (electrons per cubic centimeter) as a function of altitude in Earth's ionosphere at middle latitudes at about the period of sunspot minimum. The lowest, or D, region is not well defined so it has not been included; it is not of immediate interest because it can have no counterpart on Mars.

The E region of the terrestrial ionosphere arises mainly from ionization (the formation of electrons and positive ions) of atmospheric molecules (and some atoms) by X-rays from the Sun. In this region, at an altitude of about 90 to 150 kilometers (55 to 90 miles), the production of ions and electrons by solar X-rays reaches a maximum. In the F_1 region, between roughly 150 and 250 kilometers

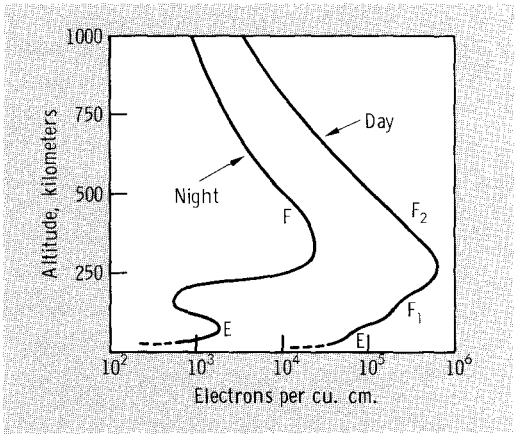


FIGURE 5.10. Electron densities in Earth's ionosphere. (After W. B. Hanson.)

(90 to 150 miles), ultraviolet radiation from the Sun causes ionization of some of the heavier species present in the atmosphere in this region; namely, molecules of oxygen and nitric oxide. Finally, above some 250 kilometers (150 miles) is the F_2 region. Here the electron (and ion) density, resulting from the ionization of oxygen atoms, which predominate at these higher altitudes, reaches a maximum and subsequently decreases steadily.

The electron density maximum in the F_2 region is attributed to the operation of two opposing factors. First, there is a tendency for the electron density to increase with increasing altitude, as indicated by the increase in density between the F_1 and F_2 regions. This is attributed to the fact that fewer electrons are being lost as a result of recombination with ions to re-form neutral atoms (or molecules). Second, at higher altitudes the mixture of electrons and relatively heavy ions, called a "plasma," diffuses downward under the influence of gravity. This plasma diffusion effect causes the electron (and ion) density to decrease, and it becomes dominant above the F_2 maximum.

Another factor which contributes to the occurrence of electron-density maxima in the ionosphere is that at higher altitudes the density of atoms and molecules in the atmosphere is so low that only a small proportion of the ionizing radiations is absorbed. This has the effect of causing the electron density to decrease with increasing altitude. At very low levels, on the other hand, very little of the radiation penetrates, and so the extent of ionization, and hence the electron density, is again low. At intermediate altitudes, however, the atmospheric density is sufficiently high to permit significant absorption of the ionizing radiation, and the electron density will tend to have a maximum value.

The decrease in the electron densities in figure 5.10 from day to night is quite marked. This is not unexpected if solar radiations are responsible for the presence of the electrons and ions in the first place. At night, electrons are lost by recombination with ions but they are not replaced by further ionization, as they would be in the daytime. The distinction between F_1 and F_2 regions also disappears at night, inasmuch as the ionization processes responsible for these regions are no longer operative.

Because Mars has an atmosphere, it was assumed that it would also have an ionosphere and scientists had speculated about its possible structure. The occultation experiment of Mariner IV provided the first definite information on the subject and, incidentally, proved most of the speculations to be incorrect. Interpretation of the ionosphere section of the phase path change curve in figure 5.8 has led to the plot given in figure 5.11 for the variation of the electron-number density with altitude above the surface of Mars. These results, obtained just prior to the occultation of Mariner IV, apply to the daytime ionosphere at 50° S latitude. They are comparable with the day curve in figure 5.10. It should be noted that the electron density scale in figure 5.10 is logarithmic, whereas in figure 5.11 it is linear.

After occultation, the radio waves from Mariner IV passed through the nighttime atmosphere of Mars and no significant electron density could be detected. This does not necessarily mean that the Martian ionosphere had disappeared completely during the night, but only that the electron density was less than 5×10^3 electrons per cubic centimeter, compared with a maximum of almost 10^5 electrons per cubic centimeter in the daytime. Consequently, a large proportion of the electrons and ions formed in the ionosphere of Mars in the daytime recombine to form neutral atoms or molecules at night, just as they do in the terrestrial ionosphere.

A possible explanation of the minor maximum at the bottom of the curve in figure 5.11 will be considered later, but first mention must be made of the discussion and differences of opinion concerning the origin of the larger maximum in the electron density. One surprising fact is that the maximum, at an alti-

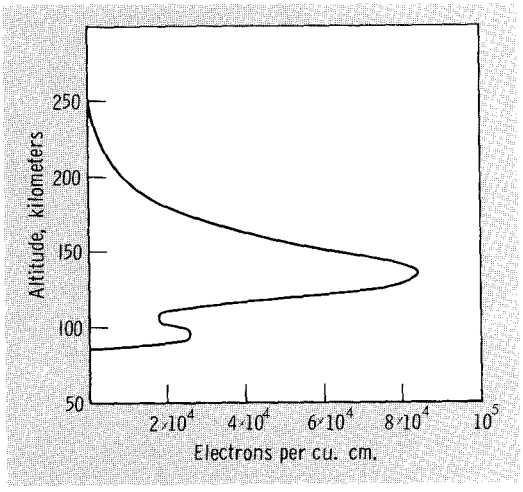


FIGURE 5.11. Electron densities in the atmosphere of Mars, derived from Mariner IV data.

tude of about 120 kilometers (75 miles), is much lower than had been expected. This indicates that the temperature of the Martian atmosphere is considerably less than anticipated. Further support for this view is provided by the slope of the electron-density curve at altitudes above the main maximum. This matter will be considered in the next section.

Attempts have been made to explain the larger electron-density maximum by associating it with either an E, an F_1, or an F_2 region. Fortunately, there are no other possibilities. The majority—but not unanimous—view seems to be that it is not an E region, but it may be either an F_1 or an F_2 region. The F_1 model of the Martian ionosphere would imply that the maximum electron density occurs at an altitude where molecules, rather than the lighter atoms, are ionized by solar ultraviolet radiation. The majority of the positive ions present would then be CO_2^+, produced by ionization of the carbon dioxide molecules which are known to be an important constituent of the atmosphere.

If the upper maximum in figure 5.11 represents an F_2 region, the predominant ion would presumably be O^+ formed by ionization of oxygen atoms by ultraviolet radiation. The oxygen atoms would themselves be produced by these radiations decomposing the carbon dioxide at lower altitudes, as described on page 81. As a result of the action of atomic and molecular diffusion, the lightest species, the oxygen atoms, would predominate over carbon dioxide and carbon monoxide at altitudes above about 80 kilometers (50 miles).

The consequences of the F_1 and F_2 models, especially with regard to the temperatures in the upper atmosphere of Mars, are quite different, as will be seen shortly. There appears to be no experimental evidence or convincing theoretical argument that can permit a choice to be made at present between the two alternative points of view. A decision must await further information that can be obtained only by the use of spacecraft.

In conclusion, some remarks may be made concerning the smaller maximum in the electron density which, as seen in figure 5.11, occurs at an altitude of about 95 kilometers. According to the calculations made in 1967 by C. Sagan and J. Veverka, this may be attributed to ionization caused by protons from the solar wind, positively charged nuclei of hydrogen with moderately high energies that are emitted continuously by the Sun. Because of Earth's magnetic field, solar-wind protons are deflected before they enter the terrestrial atmosphere so they have no effect on the ionosphere. The magnetic field of Mars, however, is very weak or nonexistent, and the solar protons would then be expected to produce a region of ionization in the Martian atmosphere, with a maximum in the electron density similar to that observed at 95 kilometers. If this maximum is indeed caused by protons, then the larger maximum at the higher altitude is expected to be an F_2 layer.

Temperatures in the Atmosphere of Mars

Numerous suggestions have been made concerning the manner in which the temperature of the Martian atmosphere varies with increasing altitude. These predictions are so very different, and based on such little reliable information, that it is not possible to summarize them here. There are, nevertheless, certain general ideas which can be described. Some are based on analogies with the known way in which the temperatures in Earth's atmosphere change with altitude, whereas others involve results derived from limited experimental measurements.

The manner in which the terrestrial atmospheric temperature varies with altitude depends on many different circumstances, such as the latitude, the season of the year, and the

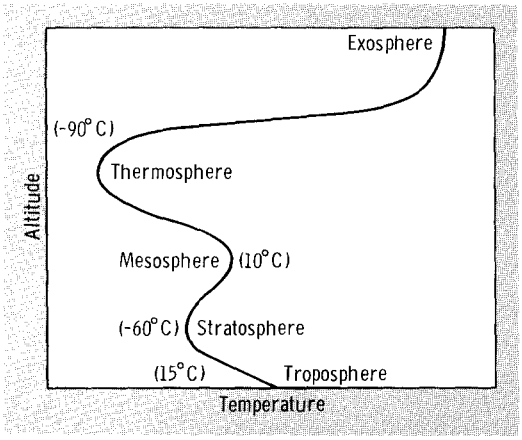


FIGURE 5.12. Temperature variations and regions in Earth's atmosphere.

activity of the Sun. In general, however, five temperature regions may be distinguished, as indicated in figure 5.12. In the lowest region, called the troposphere, the temperature decreases steadily with altitude to a height of roughly 12 kilometers (7.5 miles). The rate of decrease, referred to as the lapse rate, is almost constant at about 6.5° K per kilometer (10.5° K per mile).

The main reason the temperature decreases in the troposphere is that, as the air rises, it expands because of the decrease in the barometric pressure. The expansion is then accompanied by a lowering of the temperature. If the air is not significantly heated by radiation, the theoretical (adiabatic) lapse rate should be equal to g/c_p, where g is the gravitational acceleration and c_p is the specific heat (heat capacity per gram) at constant pressure of the air in appropriate units. For dry air, the ideal lapse rate should be 9.6° K per kilometer. The observed value is lower, partly because of the presence of water vapor, which releases its latent heat when it is condensed at higher altitudes, and partly because some heat is supplied by radiation.

Above the troposphere lies the stratosphere where the temperature remains almost constant for several kilometers and then increases with altitude to form the mesosphere, or intermediate region. The higher temperatures are caused by the absorption of solar radiation by ozone which attains its maximum concentration in this region. Above the altitude where the temperature is a maximum in the mesosphere, the amount of ozone decreases and the temperature falls correspondingly. Next, in the thermosphere, the temperature of the air is controlled by the heat taken up from the Sun and conducted along the temperature gradient. There is then a sharp increase in temperature toward a constant value which may be higher than 1000° K (727° C) in the exosphere.

In one respect, at least, the temperature variations in the atmosphere of Mars should differ from those in Earth's atmosphere. The proportion of oxygen (O_2 and O) in the Martian atmosphere is probably too small to yield any significant quantity of ozone (O_3). Consequently, the temperature will not attain a maximum in the intermediate (mesosphere) region. The general form of the temperature-altitude curve for Mars might then be like that shown in figure 5.13.

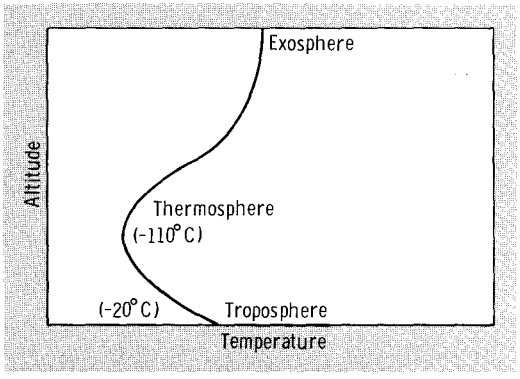


FIGURE 5.13. Approximate temperature variations and regions in the atmosphere of Mars.

This curve must not be taken too seriously, but, as far as can be determined at present, it provides a representation of the situation in broad terms. There is probably a troposphere, in which the temperature drops fairly rapidly with increasing altitude, followed by a region in which the temperature decreases more slowly. Then there is a thermosphere in which heating by solar radiation plays the dominant role, followed by an exosphere in which the atmospheric temperature attains an approximately constant value.

If the Martian troposphere consists entirely of carbon dioxide, the theoretical lapse rate (g/c_p) would be 4.7° K per kilometer (7.6° K per mile). For an atmosphere consisting of 80 percent carbon dioxide and 20 percent nitrogen, the lapse rate should be about 4.3° K per kilometer (6.9° K per mile). A rough estimate, which may be incorrect, gives the height of the Martian troposphere as about 20 kilometers. The temperature there would be, respectively, $20\times4.7=94°$ K or $20\times4.3=86°$ K, or roughly 90° K, below that at the surface of the planet. If the average temperature at the surface is taken to be about 250° K (ch. VI), the temperature at the tropopause, the top of the troposphere, would be approximately 160° K, −113° C.

Two methods have been used to estimate the average temperatures of the Martian atmosphere. One is based on the occultation measurements described earlier. According to the table on p. 90, the value was found to be about 180° K from the data before occultation and 235° K after occultation. It would appear, however, that the latter is somewhat high. The other method depends on the wavelength separations of the lines in the band spectrum of carbon dioxide in the atmosphere of Mars. These lines owe their origin to a set of specific values of the energy of rotation of the carbon dioxide molecule about one of its axes. The separations can be related to the temperature of the gas.

From the lines in the band at a wavelength close to 1.05 μ, M. J. S. Belton and D. M. Hunten of the Kitt Peak National Observatory, Ariz., in 1966 calculated a temperature of 194° K. Furthermore, they estimated that this average temperature would be equal to the actual temperature at an altitude of 1.3 times the scale height evaluated at the average temperature. The scale height, RT/Mg, with T equal to 194° K and M to 44 (for carbon dioxide), was found to be 9.3 kilometers, which is close to the Mariner IV value obtained before occultation. The average temperature of 194° K would then correspond to that at an actual altitude of $1.3\times9.3=12.1$ kilometers. If the surface temperature is 250° K, then the atmospheric temperature on Mars will have dropped $250-194=56°$ K in 12.1 kilometers. The lapse rate would thus be $56/12.1=4.6°$ K per kilometer, in good agreement with the theoretical value for carbon dioxide. Whether this result is fortuitous or not, it is not possible to say.

Another approach to the evaluation of temperatures in the upper Martian atmosphere is to utilize the electron-number density variation with altitude in figure 5.11. From the slope of the curve representing electron density in the ionosphere at levels higher than that of the maximum—above about 120 kilometers—an electron-density scale height can be calculated. This is the distance within which the electron density changes by a factor of 2.72, and it is equal numerically to $2RT/Mg$, where T is now an average temperature of the electrons and ions present and M is the molecular weight of the ions.

From the Mariner IV data, the electron-density scale height in the ionosphere at altitudes from about 120 to 250 kilometers (75 to 154 miles) was found to have a constant

value of 29 kilometers. Provided the same ions are the dominant ones throughout this region, as is probable, then the temperature would appear to be constant. Before this temperature can be calculated, the molecular weight of the ions must be known, and this is where a difficulty arises. If the electron-density maximum corresponds to that of an F_2 region, the ions should be O^+, with an atomic weight of 16. The constant temperature in the exosphere would then be about 90° K, which is extremely low. On the other hand, for the F_1 model, the ions would be CO_2^+ with a molecular weight of 44. The temperature would then be close to 250° K between 120 and 250 kilometers, at least.

In the terrestrial ionosphere the scale height data indicate that the temperature is also constant above the level of maximum electron density in the F_2 region. This level is, however, at about 400 kilometers (250 miles), and the temperature in the exosphere is over 1000° K in the daytime. If the Mariner IV scale height of 29 kilometers for the upper levels of the Martian ionosphere is correct, then the exosphere temperature would have to be very much lower, apparently not above 250° K.

Some attempts have been made to develop other models of the upper atmosphere of Mars which lead to higher temperatures in the exosphere than those derived above; for example, in the vicinity of 400° or 500° K. The basic postulates, however, appear to be incompatible with both the F_1 and F_2 models as well as with the Mariner IV value for the electron scale height. The problem of the temperatures in the Martian atmosphere must be added to the list of questions about Mars that still remain unanswered.

FOR FURTHER INFORMATION

BELTON, M. J. S.; AND HUNTEN, D. M.: The Abundance and Temperature of CO_2 in the Martian Atmosphere. Astrophys. J., vol. 145, 1966.

CHAMBERLAIN, J. W.; AND HUNTEN, D. M.: Pressure and CO_2 Content of the Martian Atmosphere—A Critical Discussion. Rev. Geophys., vol. 3, 1965, p. 299.

FJELDBO, G.; FJELDBO, W. C.; AND ESHLEMAN, V. R.: Atmosphere of Mars: Mariner IV Models Compared. Science, vol. 153, 1966, p. 1518.

KAPLAN, L. D.; MÜNCH, G.; AND SPINRAD, H.: An Analysis of the Spectrum of Mars. Astrophys. J., vol. 139, 1964, p. 1.

KELLOGG, W. W.; AND SAGAN, C.: The Atmospheres of Mars and Venus. Nat. Acad. Sci.-Nat. Res. Council Publ. 944, 1961.

MCELROY, M. B.: The Upper Atmosphere of Mars. Astrophys. J., vol. 150, 1967, p. 1125.

MICHAUX, C. M.: Handbook of the Physical Properties of the Planet Mars. NASA SP–3030, 1967.

MOROZ, V. I.: Physics of Planets. Ch. II, NASA TT F–515, 1968.

ÖPIK, J. E.: Atmosphere and Surface Properties of Mars and Venus. Prog. Astronaut. Sci., vol. 1, 1962, p. 261.

OWEN, T. A.: The Composition and Surface Pressure of the Martian Atmosphere: Results from the 1965 Opposition. Astrophys. J., vol. 146, 1966, p. 257.

SAGAN, C.; HANST, P. L.; AND YOUNG, A. T.: Nitrogen Oxides on Mars. Planetary and Space Sci., vol. 13, 1965, p. 73.

SAGAN, C.; AND VEVERKA, J.: Martian Ionosphere: A Component Due to Solar Protons. Science, vol. 158, 1967, p. 110.

SCHORN, R. A.; ET AL.: High-Dispersion Spectroscopic Observations of Mars: II, The Water-Vapor Variations. Astrophys. J., vol. 147, 1967.

SLIPHER, E. C.: The Photographic Story of Mars. Sky Publishing Corp., 1962.

SPINRAD, H.; ET AL.: High-Dispersion Spectroscopic Observations of Mars: I, The CO_2 Content and Surface Pressure. Astrophys. J., vol. 146, 1966, p. 331.

VAUCOULEURS, G. DE: Physics of the Planet Mars. Faber & Faber, 1954.

VI

The Surface of Mars

THE POLAR CAPS

General Description of the Polar Caps

THE MAIN ASPECTS OF THE MARTIAN SURFACE were described briefly in chapter I. They will now be considered in greater detail, starting with the polar caps. During the course of the local autumn and winter on Mars, a brilliant white cap, steadily increasing in size, develops over each of the poles in turn. Most of the time the polar cap appears to be covered by a white cloud or hood, but toward the end of the winter, just before the local spring equinox, the cloud disappears and the sharp outlines of the cap itself become clearly visible. Subsequently, during the spring, the size of the cap gradually decreases and reaches a minimum shortly after the summer solstice, around the beginning of the local summer.

Since winter in the northern hemisphere of Mars corresponds to summer in the southern hemisphere, the northern polar cap grows at the same time as the area covered by the southern cap diminishes. Half a Martian year later, the situation is reversed; the cap around the north pole gets gradually smaller, during the local summer, while the southern cap grows in size. This seasonal waxing and waning, in turn, of the polar caps was reported by William Herschel at the end of the 18th century, and it has been observed repeatedly since that time. It is only at certain Martian apparitions that both caps can be seen simultaneously, of course, and these conclusions have been drawn mainly from observations on each cap made at different times.

When it reaches its maximum size, the southern polar cap covers a larger area than does the northern cap. The reason is that the autumn and winter, when the cap is being formed, is longer (and colder) in the southern hemisphere (381 Earth days) of Mars than in the northern hemisphere (306 days). The northern polar cap does not often extend farther south than a latitude of 60° N, but the southern cap generally reaches as far as 50° S (fig. 6.1), and sometimes to 45° (or less) S latitude. It is of interest that the extent of the polar caps on Mars, relative to the size of the planet, is much the same as on Earth. The terrestrial polar caps, as indicated by the snow cover at high altitudes, extend to about 50° latitude by the end of winter.

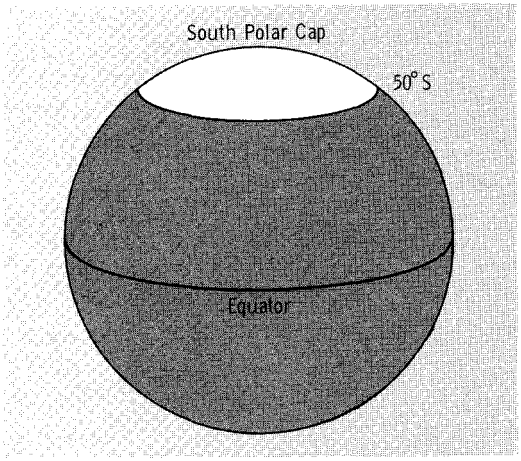


FIGURE 6.1. Approximate representation of the maximum extent of the south polar cap of Mars.

In spite of its larger area at the end of the local winter, the south polar cap on Mars decreases in size much faster than does the cap in the northern hemisphere. This is so because the summer, although shorter than in the north, is much hotter in the southern hemisphere, since the planet is then nearer to the Sun. In some years, the cap around the south pole vanishes completely. The cap in the northern hemisphere, on the other hand, may become quite small, about 320 kilometers (200 miles) across, but it has never been known to disappear entirely.

The remnants of the polar caps are not centered over the actual north and south poles; that is, the points through which the axis of rotation of Mars passes. The locations are, however, always the same. The center of what remains of the south polar cap at the end of the local summer is some 400 kilometers (250 miles) from the south pole, whereas in the northern hemisphere the distance from the center of the cap to the pole is only about 65 kilometers (40 miles).

The deviations of the center of the polar caps from the poles themselves, which are expected to be the coldest places, are commonly attributed to differences in altitude of the surface in the polar regions. It has been assumed that there are elevated regions centered at distances of 400 kilometers from the south pole and 65 kilometers from the north pole. By analogy with conditions on Earth, it is thought that these elevated regions would warm up more slowly than the lower areas around the poles themselves. As will be seen shortly, however, some scientists dispute this conclusion.

Recession of the Polar Caps

Apart from relatively small variations from year to year, such as might occur on Earth, the rate at which the polar caps recede is remarkably regular. E. C. Slipher of the Lowell Observatory, Ariz., has compiled the results of observations of the south polar cap made over a period of several months at the times of a number of appropriate oppositions in 1798 and from 1877 through 1924. Furthermore, he checked the measurements with various drawings made between 1781 and 1862, and found them to be in general agreement. In figure 6.2, the dimensions of the cap, as indicated by its width and by the

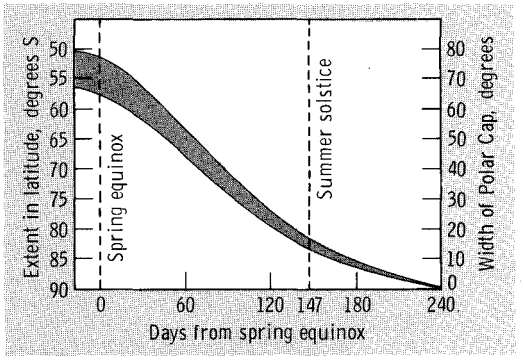


FIGURE 6.2. Variation of dimensions of the south polar cap during the Martian spring and summer. (After E. C. Slipher.)

latitude to which it extends, are plotted against the number of Earth days after the Martian spring equinox in the southern hemisphere. The breadth of the curve is a measure of the deviations in size of the polar cap from year to year and of the results obtained by different observers in any given year. Moreover, because of local differences in the surface elevations, the extent of the cap is not the same at all longitudes.

On the whole, Slipher concluded that the "study revealed no evidence of any irregularity in the . . . [recession] of the south polar cap . . . during this long period of observations." Although such detailed measurements are not available for the north polar cap, the indications are that it decreases in size at about the same rate every year.

It is probable that the polar caps also grow in size in a regular manner during the local autumn and winter, but it has not been possible to verify this by actual measurements. While the caps are increasing in size they are generally covered by white clouds, as noted earlier, and so the caps themselves are rarely visible. It is only just before the spring equinox that the cloud, or hood, begins to lift. The sharp outlines of the polar cap then appear and the cap itself is seen to be somewhat smaller than the hood. Thus, the actual size of the cap is not known during its period of growth. It is only from a few days before the local spring equinox until about 90 days after the summer solstice (fig. 6.2) that measurements can be made of the extent of the polar cap.

When a polar cap starts to recede, its outer rim becomes ragged. Some parts are seen to extend much farther than others, leaving projections which may later become detached from the main polar cap. About the middle of spring, dark rifts appear at various locations in the cap. The rifts increase in size and cause the cap to break up into a number of isolated regions which gradually decrease in area (fig. 6.3). Conversely, during the winter, these regions are the ones that form first as the polar cap develops.

"Each Martian year," says A. Dollfus, "the same details are seen in the same locations during the same part of the Martian season . . . although some delays or variations may

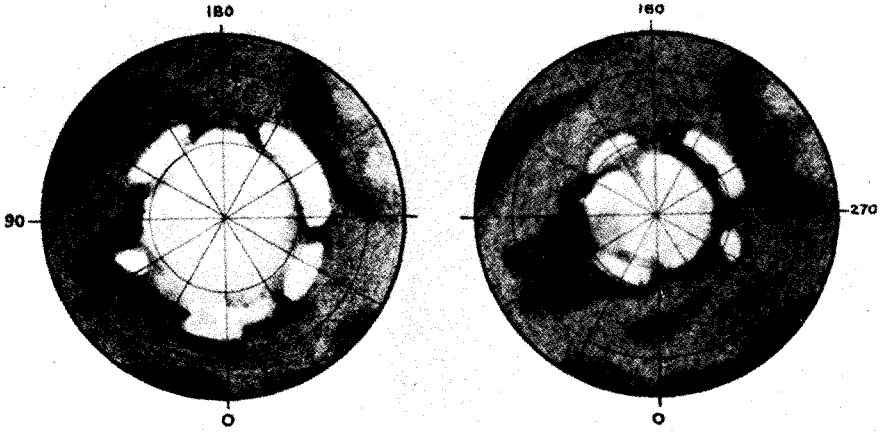


FIGURE 6.3. Changes in the polar cap as it recedes in the spring. (Drawings by A. Dollfus; courtesy University of Chicago Press.)

occur, as is characteristic of natural phenomena." The irregular breakup of the polar caps, which follows the same pattern from year to year, must be attributed to permanent local conditions, possibly variations in altitude.

One of the best known of the polar-cap irregularities, which has been observed repeatedly during the recession of the southern cap, consists of some exceptionally bright isolated spots. They were discovered by O. M. Mitchel, at the Cincinnati Observatory, in 1845, and were named the Mitchel Mountains (or Mountains of Mitchel) by N. Green in 1877. These "mountains" appear regularly at the same place (latitude 73° S, longitude around 290° W) and they are always seen as isolated spots on about the same date in the Martian year (fig. 6.4). In general, the bright spots vanish after a few days and only rarely can they be seen for as long as 2 weeks.

In 1894, Percival Lowell reported observing the Mountains of Mitchel some days earlier than usual, while they were still surrounded by parts of the polar cap. "As I was watching the planet," he wrote, "I saw suddenly two points flash out in the midst of the polar cap. Dazzlingly bright upon the duller white background . . . these stars shone for a moment and then slowly disappeared." The estimated latitude and longitude of the bright spots indicated that they were two of the Mountains of Mitchel.

It has usually been assumed that the isolated parts of the polar caps, such as the Mountains of Mitchel, which remain after the surrounding areas of the cap have disappeared, are elevated regions. It is true that this would generally be the case on Earth, where snow remains on mountain peaks long after it has gone from adjacent areas at lower altitude. There is, however, no evidence that this is so on Mars. It will be seen in due course that under certain circumstances the

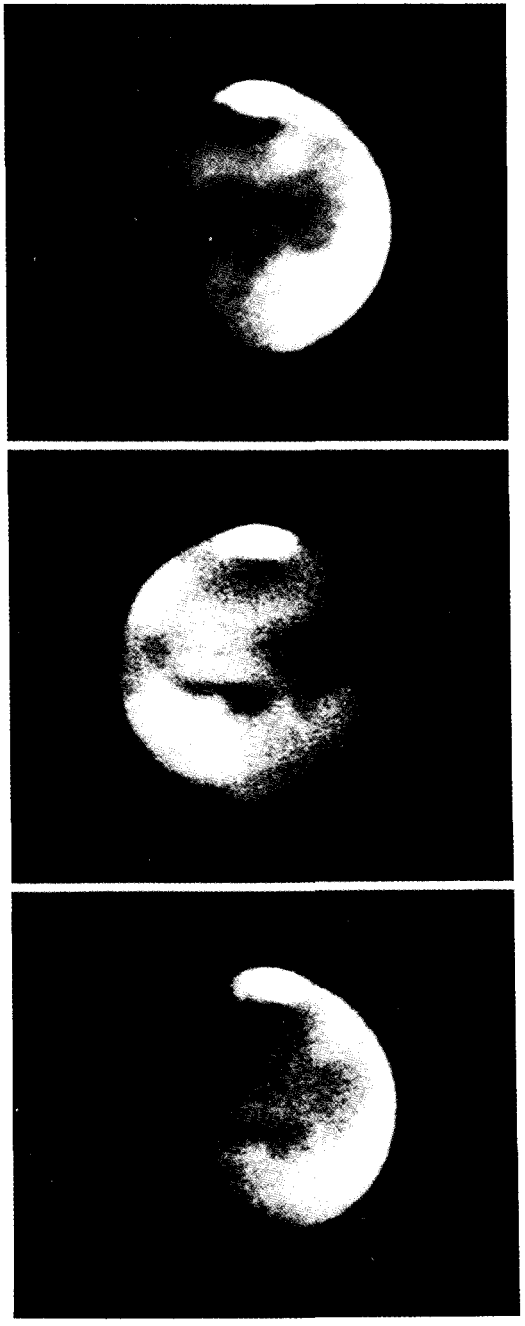


FIGURE 6.4. The Mountains of Mitchel in 1909, 1924, and 1941. (Lowell Observatory photographs.)

material forming the polar caps may possibly disappear first from the higher altitudes.

As each polar cap recedes, a dark band, or fringe, appears to form at the edge. The band follows—"tightly hugs" in the words of E. C. Slipher—the polar cap during its recession. At first, the band is fairly wide, but it shrinks as the polar cap decreases in size. After the summer solstice, when the cap is quite small (fig. 6.2), it is described as being a "barely discernible thread" around the white cap. On the occasions when the south polar cap disappears completely, so also does the dark band.

A dark band was evidently seen by W. Beer and J. H. von Mädler around the north polar cap in 1830, but it attracted little or no interest until P. Lowell observed the formation of such a band in the southern hemisphere of Mars in 1894. Since that time, the dark bands around the polar caps have been reported on several occasions. Some astronomers have claimed that the apparent band is an optical illusion, resulting from the contrast between the bright polar cap and the surrounding darker areas, but the evidence, on the whole, seems to point to its reality as a Martian phenomenon.

Writing in 1906, Lowell described the band as being "deep blue" in color and referred to it as a "badge of blue ribbon about the melting cap." There is some doubt, however, about the blue color, and most later observers have said that under good seeing conditions it appears to be dark brown or black. According to A. Dollfus, small telescopes show a continuous dark belt surrounding the polar cap, but at greater magnification the band is seen to consist of many separate dark spots.

Composition of the Polar Caps

Until recent years, most astronomers accepted the assumption made by William Herschel that the Martian polar caps, like those of Earth, consisted of some form of solidified water. In the early 1890's, however, A. C. Ranyard and G. J. Stoney, apparently independently, had suggested that the polar caps might be made up of solid carbon dioxide, a white crystalline substance similar to snow in its general appearance.[1] At the time there was no evidence that the Martian atmosphere contained substantial quantities of carbon dioxide gas and the possibility was not taken seriously. In 1954, G. de Vaucouleurs wrote: "The polar caps are without the slightest doubt layers of crystallized water—probably more like white frost than solid ice or snow."

In 1895, P. Lowell had suggested that the water in the Martian polar caps is "probably deposited as hoarfrost"; that is to say, the material is deposited directly from the atmospheric water vapor as a solid without the intermediate formation of liquid water. If the polar caps are actually a form of solidified water, then it is certain that they must be hoarfrost deposits. If the amount of water vapor in any atmosphere, either of Earth or Mars, contributes a pressure, i.e., it has a partial pressure, of less than about 6 millibars, hoarfrost, but not liquid water, will form when the temperature falls sufficiently below 0° C.

The average vapor pressure of water in the Martian atmosphere is much less than 6 millibars. In fact, the total pressure of all the atmospheric constituents, of which water vapor is a very minor one, may not be greatly in excess of 6 millibars. If solid water is condensed from atmospheric water vapor on a cold area of Mars, it must be in the form of hoarfrost. Conversely, when the planet warms up, it is expected that the solid hoarfrost will be converted directly into vapor, a process called sublimation.

[1] The commercial material known as "Dry Ice" is made by compressing solid carbon dioxide "snow."

319-889 O—69——4

The first report concerning an experimental study of the composition of the Martian polar caps was made by G. P. Kuiper in 1948. He compared the infrared reflection spectrum from the north polar cap with the spectrum of ordinary snow, on the one hand, and with carbon dioxide snow, on the other. Snow, like hoarfrost, is produced by direct condensation from the vapor to the solid state. The difference between hoarfrost and snow is that the former is deposited on a cold surface, whereas the latter is produced in the atmosphere at high altitude and then falls to the ground.

Kuiper found that the spectrum of ordinary (water) snow was similar to that of the Martian polar cap, but the spectrum of carbon dioxide snow was quite different. He stated, therefore, "that the Martian polar caps are not composed of carbon dioxide and are certainly composed of water frost at low temperature [much below 0° C]." From a study of the infrared reflection spectra, published in 1966, V. I. Moroz in the U.S.S.R. also concluded that the polar caps consist of solid water, rather than of carbon dioxide.

Apparent confirmation of this view was reported by A. Dollfus in 1950 on the basis of the polarization (p. 85) of the scattered light reflected from the polar caps. He showed that, when hoarfrost is heated by an electric arc, it partly sublimes into vapor, without the intermediate formation of liquid water. "The remainder," wrote Dollfus, "takes on the appearance of opal glass, full of small holes and cavities . . . and the polarization becomes very similar to that of the Martian polar caps . . . It thus seems probable that the white spots [caps] at the poles are deposits of hoarfrost." Unfortunately, Dollfus did not perform analogous experiments with solid carbon dioxide for comparison.

The albedo (p. 65) of a thick layer of fresh snow is 0.8, but the value found for the polar caps of Mars is only about 0.5. In the experiment referred to above, Dollfus noted that the porous material remaining after partial sublimation of hoarfrost had a lower albedo than originally. Thus, the surface reflecting power of the polar cap might well depend on its condition, such as size of crystals, presence of small cavities, etc. This dependence of reflectivity on particle size is well known, too, for materials other than hoarfrost.

By comparing the rate of recession of the polar caps during the spring with the heat absorbed from the Sun, G. de Vaucouleurs concluded that the average thickness of the solid layer of hoarfrost is "of the order of a few centimeters (about 1 or 2 inches)." He went on to say: "Owing to the thinness of the layer, it seems likely that it does not cover completely and uniformly . . . the roughest parts of the polar areas. This lack of continuity easily accounts for the small value of the apparent albedo." The thinness of the caps would also account for their relatively rapid disappearances during the local spring. On Earth, the polar caps are much thicker. In some regions near the poles, they are measured in miles rather than in inches.

If the polar caps consist of a form of solidified water, then their growth and regression may be supposed to occur in the following manner. When the surface temperature in the vicinity of one of the Martian poles falls sufficiently below 0° C (probably to –80° C) in the local autumn, hoarfrost condenses out from the water vapor in the atmosphere. The polar cap then starts to develop and to increase in size. As the autumn proceeds and passes into winter, the area where the temperature is always considerably below 0° C, during both day and night, extends farther and farther from the pole to successively lower latitudes. Thus, the polar cap grows steadily in size.

Toward the end of winter, the edge of the cap has reached a latitude where either the

temperature is not low enough for hoarfrost to form or where the hoarfrost deposited during the Martian night disappears as the surface is warmed by the Sun in the daytime. The polar cap has then attained its maximum size.

As the winter ends and spring begins, the surface of the planet gets warmer and the hoarfrost starts to sublime (turn directly into water vapor). The temperature will first increase at low latitudes, in the regions farthest from the pole, and so the polar cap will start to recede. Apart from variations resulting from special local conditions, possibly differences in elevation, the size of the cap will decrease throughout the spring and early summer. Apparently, in some years the temperatures near the south pole rise sufficiently to permit the cap to disappear. This is not the case, however, in the northern hemisphere, with its cooler summer. The north polar cap diminishes in size, but it has never been observed to disappear completely.

One polar cap on Mars is always growing while the other is receding. In view of the small total quantity of water vapor in the atmosphere, it is evident that, assuming the caps to consist of hoarfrost, there is a continuous transfer of water vapor back and forth across the planet from one hemisphere, where the polar cap is receding in the local spring and summer, to the other hemisphere, where the cap is advancing in the autumn and winter. There is evidence from the seasonal changes in abundance of water vapor in the atmosphere of the two hemispheres, mentioned in chapter V, that such a transfer does take place.

The dark band (or collar) that surrounds each polar cap as it recedes has been ascribed to moist ground. This would imply the formation of liquid water at the edge of the cap. If equilibrium exists between the hoarfrost of the polar cap and the water vapor in the atmosphere, the hoarfrost would not be converted into liquid water. But it is not impossible that where the temperature of the ground at the periphery of the polar cap exceeds 0° C, local conditions may permit the formation of some liquid water. As the atmosphere warms up during the summer, the water would evaporate and the dark collar would disappear, as is actually observed.

According to A. Dollfus, the polarization of the dark band does not agree with that of moist soil, but because polarization depends on many factors, it is not in this instance a compelling argument, one way or the other. There is, however, no reasonably satisfactory alternative for the not too probable explanation that the dark collar to the receding polar cap is caused by moist soil.

Carbon Dioxide in the Polar Caps

In spite of the evidence which appears to indicate that the polar caps of Mars consist of solidified water, several scientists during the early 1960's have revived the view that carbon dioxide is the sole (or chief) constituent. It is now known that the Martian atmosphere is composed mainly of carbon dioxide gas, and that its abundance is much larger than that of water vapor. In fact, the total mass of carbon dioxide in the atmosphere of Mars is probably some 10 000 times as great as that of water vapor. The problem of the transfer of carbon dioxide back and forth from one hemisphere to the other, as the polar caps wax and wane in turn, would thus be much less severe than for water vapor. The idea that the polar caps may be made up of solid carbon dioxide, rather than of solidified water, appears to be less improbable than it did to P. Lowell, who in 1895 referred to its proponents as having "that class of mind which likes to make of molehills of questions, mountains of doubts."

If the Martian polar caps consist largely

of solid carbon dioxide, then the latter would be deposited, like hoarfrost, directly from the gas present in the atmosphere, without the intermediate formation of liquid, when the surface of the planet becomes cold enough. Similarly, upon warming up, during the spring, the solid carbon dioxide would sublime and be reconverted into gas. The transfer of the gas from one hemisphere to the other would probably be accompanied by a change in the carbon dioxide abundance in the atmosphere. By 1968, however, no measurements had been made whereby such a change might be observed.

One of the arguments against the view that the polar caps are solid carbon dioxide is that the temperatures attained are not low enough for the carbon dioxide gas in the atmosphere to condense out as a solid. The relatively limited, and not too accurate, estimates of surface temperature, based on measurement of the infrared radiation from Mars indicate that the average daily (day and night) temperature in the polar regions never gets below about −100° C (173° K). This would certainly not be low enough to permit the deposition of solid carbon dioxide unless the pressure of the gas in the Martian atmosphere were much higher than it is thought to be. The partial pressure of carbon dioxide in the atmosphere of Mars is about 5 to 7 millibars (see p. 90), and it is known from laboratory measurements that, in these circumstances, the temperature would have to be less than approximately −125° C (148° K) for condensation of solid carbon dioxide to occur.

In 1966, R. B. Leighton and B. C. Murray of the California Institute of Technology reported results of calculations concerning the surface temperature at various latitudes during the course of a Martian year. These calculations, which were admittedly oversimplified, took into consideration the heat absorbed by the surface from solar radiation, the heat emitted from the surface, and the heat exchanged between the surface and the underlying layers of the ground by thermal conduction. Two important results were obtained in favor of the view that carbon dioxide could be condensed in solid form from the atmosphere of Mars.

First, at latitudes above about 50° N and 45° S, respectively, the computations indicated that the temperature should fall low enough to permit solid carbon dioxide to be formed. If allowance is made for the requirement that the temperature must remain below −125° C (or so) during the day as well as the night, these calculated latitudes are in good agreement with the average latitudes (about 60° N and 50° S) to which the polar caps are observed to extend.

Second, by taking into consideration the latent heat of sublimation, which must be provided when either solid carbon dioxide or solidified water is converted into vapor, the theoretical rate at which the polar cap should recede during the spring and early summer can be determined. It was found that if the caps are assumed to be solid carbon dioxide, the calculated rate agrees well with that observed (fig. 6.2). On the other hand, if water is the main component of the polar caps, the calculated and observed recession rates are quite different.

The maximum amount of solid carbon dioxide which might be expected to condense on the polar caps is estimated to be from 100 to 150 g/cm^2. The density of solid carbon dioxide in closely packed form is 1.56 g/cm^3. Such a deposit would represent a minimum thickness of from 58 to 96 centimeters (roughly 2 to 3 feet). Since it is unlikely that the carbon dioxide crystals will be tightly packed, the average thickness of the polar caps might be 3 to 5 feet.

Another approach to the subject was made independently in 1966 by C. Leovy of the

National Center for Atmospheric Research, Boulder, Colo. During the period of winter darkness in the polar regions, essentially no heat is received from the Sun and conduction in the ground can be ignored. There is then a balance between the heat radiated from the Martian surface into the atmosphere and that returned to the surface from the atmosphere by radiation and by turbulent motion. This balance involves the temperatures of both the atmosphere and the ground, as well as certain characteristic properties of the atmosphere.

By combining available experimental data with the heat balance equation, Leovy calculated the minimum atmospheric temperature that would prevent the Martian surface from reaching −125° C when solid carbon dioxide can be deposited. The results showed that, for a clear atmosphere, meaning one free from clouds, it is probable that the condensation of solid carbon dioxide will occur in the winter polar region. On the other hand, if the local atmosphere contains a cloud of microscopic ice crystals, with a total mass of 0.001 g/cm^2, then the deposition of solid carbon dioxide on the ground might be partially inhibited but not prevented. Because such a cloud would require almost all the water present in the Martian atmosphere, it is very unlikely that it could ever form over the winter pole. The conclusion drawn was "that carbon dioxide condensation in the Mars polar caps is quite likely. However, the deposition rate may be strongly dependent on the formation of extremely tenuous water ice clouds in the winter polar regions."

Further evidence that the polar caps might consist of carbon dioxide was reported in 1968 by D. M. Morrison and C. Sagan. They used infrared measurements, made by W. H. Sinton and J. Strong in connection with the temperature measurements described on page 133, to calculate the temperature gradient on the Martian surface between the equator and 45° latitude. A relatively short extrapolation to the latitude of the edge of the polar cap gave a predicted surface temperature that is consistent with the formation of solid carbon dioxide.

If the polar caps of Mars are made up mainly of solid carbon dioxide, then what of the infrared reflection spectra and the polarization measurements which seem to indicate that the polar cap material is a form of solidified water? In answer to this question, Leighton and Murray have stated: "We are not aware of a sufficiently thorough . . . study of the reflection spectra of both solid water and carbon dioxide under simulated Martian conditions to justify the identification." Furthermore, they point out that Dollfus "found difficulty in producing a form of [solidified] water . . . that would exhibit the polarization . . . [of the Martian polar caps] . . . and he did not study the properties of carbon dioxide frost." It would appear, therefore, that the available evidence from reflection spectra and polarization is not conclusive. More extensive experimental studies in this area are clearly desirable.

Even if the polar caps are largely carbon dioxide, it does not mean that they do not also contain some solidified water. Because hoarfrost will condense from water vapor in the Martian atmosphere around −80° C, before the temperature gets low enough for solid carbon dioxide to form, there must inevitably be some solid water in the polar caps. The observed variations in the water-vapor content of the atmosphere according to the local season provides support for this view. In the local spring, as the surface warms up, the carbon dioxide would sublime more readily than the hoarfrost, and there would be a tendency for solidified water to remain.

A consideration of the atmospheric temperatures derived from the Mariner IV occultation experiment described in chapter V has

led G. Fjeldbo, W. C. Fjeldbo, and V. R. Eshleman of the Stanford University Center for Radar Astronomy to the following conclusions:

> The low daytime temperature in the lower atmosphere over Electris tends to support the carbon dioxide polar-cap theory. . . . The winter polar cap may be predominantly dry ice [i.e., solid carbon dioxide], while the bottom part of the thicker central region, some of which persists through the summer, may be largely water ice.

The formation of liquid carbon dioxide from the solid is extremely improbable under any conditions likely to exist on Mars. Ground moistened by liquid carbon dioxide could not possibly account for the dark band which hugs the polar cap as it recedes in the local spring. It might be produced by liquid water if the polar cap contains a sufficient quantity of solidified water in addition to carbon dioxide. On the other hand, it is quite possible that the dark collar around the polar cap is not caused by moist soil but by some other, but still unknown, circumstances.

The greater the partial pressure of carbon dioxide gas in the Martian atmosphere, the higher would be the temperature at which solid carbon dioxide would condense out. This is a general conclusion that is known from laboratory measurements and is applicable under all circumstances. Because the atmosphere of Mars appears to consist mainly of carbon dioxide, the higher the atmospheric pressure, the higher the temperature at which solid carbon dioxide will deposit on the surface when its temperature drops. In other words, solid carbon dioxide will condense more readily in regions where the atmospheric pressure is high. Conversely, as the ground warms up, solid carbon dioxide will tend to vaporize most readily from areas where the atmospheric pressure is low, and least readily where the pressure is high.

The surprising conclusion to be drawn from the foregoing arguments is that solid carbon dioxide may be expected to form earliest in the autumn, and to disappear latest in the summer, at low altitudes where the barometric (and carbon dioxide) pressure is highest (fig. 6.5). This is quite opposite to the long-accepted view based on the terrestrial analogy, that the parts of the polar caps at the highest altitudes remain while the surroundings at lower altitudes disappear.

These considerations led B. T. O'Leary and D. G. Rea, University of California, Berkeley, in 1967, to suggest that the temporary isolated bright patches in the polar caps, observed repeatedly at the same locations, are actually depressions, such as craters or valleys, rather than elevated areas. A similar conclusion was reached independently by J. B. Pollack and C. Sagan on somewhat different grounds (p. 122). If the bright areas are actually at lower elevations than their surroundings then the supposed Mountains of Mitchel should more properly be named the Depressions of Mitchel. The permanent north polar cap, centered some 65 kilometers from the pole, has been attributed to the presence of a plateau at an elevation of about 1000 meters (3300 feet). If the views ex-

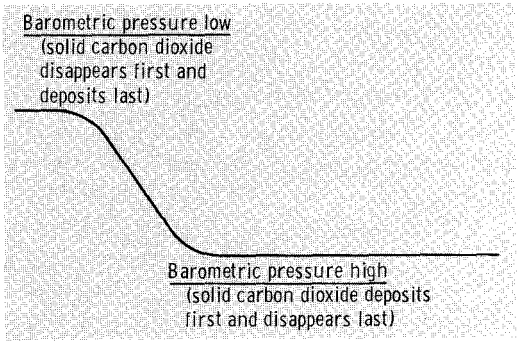


FIGURE 6.5. Deposition and disappearance of solid carbon dioxide at different elevations.

pressed above are correct, this region would be a large, flat valley rather than a plateau.

Nitrogen Dioxide and the Polar Caps

Brief mention may be made of another theory of the nature of the Martian polar caps. It will be recalled from chapter V that, in 1960, C. C. Kiess and his associates had put forward the idea that the Martian atmosphere contains significant quantities of nitrogen dioxide (NO_2) gas. Molecules of this gas can combine in pairs, especially at moderately low temperatures, to form the pale-yellow gas nitrogen tetroxide (N_2O_4). At normal terrestrial atmospheric pressure the latter condenses at $-9°$ C to a whitish solid. The suggestion was made that the polar caps are composed mainly of solid nitrogen tetroxide. The chief obstacle to the acceptance of this view is that the quantity of nitrogen dioxide in the atmosphere of Mars appears to be insignificant (p. 81).

THE BRIGHT AND DARK AREAS

The Bright Areas

The orange and yellowish-brown areas which constitute the major part—about 70 percent—of the surface of Mars, mainly in the northern hemisphere, are responsible for the reddish color of the planet as seen by the unaided eye. In 1809, H. Flaugergues thought that the reddish areas were large clouds, but their permanent nature indicated to others that this was not so. Later, they were regarded as dry land, by contrast with the dark areas which were considered to be bodies of water, and the brighter areas were commonly known as continents. Because of their appearance, W. H. Pickering in 1886 referred to the orange and brown regions as deserts, and this name has been commonly used, although the areas probably bear little, if any, resemblance to terrestrial deserts.

Since the words "continent" and "desert" have certain implications, which are undoubtedly incorrect, it is preferable to avoid their use entirely. The regions under consideration are best described by the noncommittal term bright areas. They do not all have exactly the same color, but reds and yellows predominate in all cases. The adjective "bright" is used only to indicate brightness relative to the dark areas. Thus, the albedo in red light, for which the value is largest in the visible spectrum, is only 0.28 for the bright areas. This is relatively high in comparison with 0.16 for the dark regions. As noted earlier, the albedo for the polar caps is about 0.5.

In 1934, the German-born astronomer Rupert Wildt suggested that the bright areas on Mars were composed "of strongly oxidized sandy formations, with iron almost completely in the form of the oxide Fe_2O_3 [ferric oxide]." This oxide exists in different forms in several terrestrial minerals, but they are all characterized by their red, brown, and yellow colors.

An attempt at a more precise characterization of the oxide was reported by A. Dollfus in 1951 and confirmed in 1957 on the basis of polarization measurements made during the four Martian apparitions from 1948 through 1954. He compared the curve showing the polarization of the bright areas on Mars as a function of phase angle (p. 43) with that of various terrestrial minerals, and concluded that "the polarization of the desert regions on Mars is especially well reproduced by limonite, a hydrated iron [ferric] oxide . . . in a finely pulverized condition." A comparison of the polarization of the bright areas on Mars (full curve) with that of powdered limonite (dots) is given in figure 6.6.

Observations of the color and albedo of Mars, reported by V. V. Sharonov in the

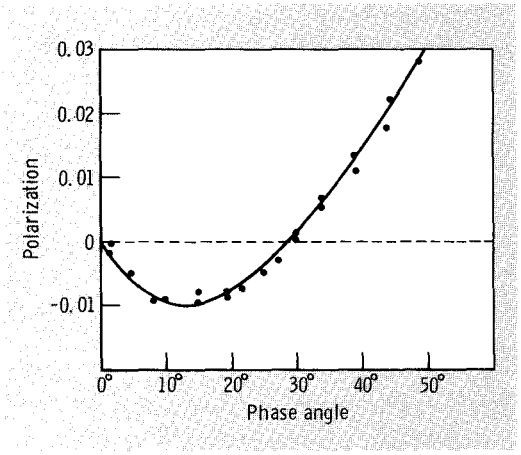


FIGURE 6.6. Comparison of polarization of the bright areas on Mars (full line) with that of powdered limonite (dots). (After A. Dollfus.)

U.S.S.R. in 1961, and of the infrared reflection spectra, by C. Sagan and his collaborators in 1965, provided support for the view that the bright areas are covered by the mineral limonite, which is so named because of its lemon-yellow color. It has been pointed out, however, that limonite is not a well-defined material and that the word is used in geology to include several ferric oxide minerals of different crystalline forms and containing different amounts of water. Disregarding impurities, the general composition may be represented by $Fe_2O_3 \cdot xH_2O$, where x is a variable quantity. Thus, in 1964, A. L. Draper and his coworkers were able to reproduce the infrared reflection spectrum of Mars in the wavelength region of 1 to 2 μ (p. 76) by means of a finely powdered mixture of the minerals goethite ($Fe_2O_3 \cdot H_2O$) and hematite (Fe_2O_3). The latter is so named for its blood-red color.

A detailed review of the infrared spectra and polarization measurements on Mars and on various ferric oxide minerals has led J. B. Pollack and C. Sagan to conclude in 1967 that the bright areas of Mars are covered with significant quantities of powdered limonite, consisting mainly of goethite, with a minor proportion, if any, of hematite. There must consequently be at least one molecule of water to each molecule of ferric oxide. The fact that the molecule of water in goethite is apparently water of hydration, which can be driven off fairly readily by heat, is important in relation to the question of the availability of water on Mars (ch. X). The average diameter of the limonite particles was estimated to be about 25 microns.

Regardless of the exact significance of the term "limonite," it seems to be generally agreed that the color of the bright areas on Mars is attributable to the presence of hydrated ferric oxide (or oxides). There are differences of opinion, however, concerning the relative abundance of this material. Some scientists claim that it is a major component of the bright areas, at least of the surface layers, whereas others argue that this is not so.

The suggestion made in 1964 by A. B. Binder and D. P. Cruikshank is that the limonite is merely a fairly thin surface stain formed by oxidation and hydration, the action of oxygen and water, respectively, on iron-bearing minerals in igneous rocks formed in the interior of the planet. This view is based on the observation that the infrared reflection spectrum of Mars can be well matched by rocks of this kind found in deserts on Earth. If the Martian surface consists of such a material, then there must have been a time when the atmosphere of the planet contained substantial quantities of oxygen and water vapor. Incidentally, surface temperature measurements, to be described later in this chapter, show that the bright areas of Mars must be covered mainly by a finely powdered material rather than by rocks of substantial size.

Another possibility, based on geological

considerations, has been proposed by R. A. Van Tassel and J. W. Salisbury in 1964. They suggested that powdered silicate minerals, formed from pulverized crustal rock, comprise the major portion of the surface of Mars, just as they do on Earth and evidently also on the Moon, as indicated by data from the Surveyor spacecraft. Furthermore, the composition of stony meteorites called chondrites indicates that there should be relatively more silicon than iron in the outer layers of Mars.

An added complication is provided by W. M. Sinton's contention that the strong emission band in the spectrum of Mars at a wavelength of 3.1 μ indicates the presence of hydrated minerals on the surface of the planet. Sinton and others, however, have called attention to the weakness of the band at about 0.88 μ normally exhibited by limonite. This implies that limonite cannot be the major hydrated mineral on the Martian surface. It is presumably not a silicate because silicates in nature are formed at high temperatures and do not contain water of hydration.

If the proportion of iron on the Martian surface is actually high, there are two factors which could be responsible. First, if there has been little or no differentiation in the interior of Mars, as suggested in chapter IV, the density of the outer layers, including the surface, will be approximately the same as the average value for the planet, about 4 g/cm^3. It is to be expected that the surface material on Mars will contain a larger proportion of heavy elements, such as iron, than is present on Earth, in which differentiation has occurred. The density of the surface rocks on Earth ranges from about 1.6 to 3.5 g/cm^3. It is probably a coincidence that the density of limonite is close to that of the average density of Mars as a whole.

A further point to bear in mind is that Mars is close to the asteroidal belt (ch. III). For this reason, meteorites, many of which contain considerable proportions of iron, may have fallen on Mars over the ages in greater numbers than on Earth or the Moon. This possible way of accounting for the large proportion of iron on the Martian surface was apparently first proposed by the famous Swedish chemist Svante Arrhenius in 1910, and it has been revived in recent years. The oxygen required to produce the highly oxidized ferric state, in which the iron now appears to exist on the surface of Mars, would presumably have been derived from water by the action of ultraviolet radiation from the Sun.

The Dark Areas

For many years astronomers, including G. V. Schiaparelli, thought that the darker areas on Mars were actually bodies of water, and they were called maria (plural of the Latin *mare,* meaning sea). In 1892, W. H. Pickering claimed to have detected some canals crossing the mare, and this observation was confirmed in 1894 by A. E. Douglass. The idea which then came into prominence, based on a suggestion originally made by A. Liais in 1860, was that the darker areas on Mars were covered with vegetation. Some students of the planet still favor this point of view, but there are other possible explanations for these areas. In any event, they are certainly not seas, and they will be referred to here simply as the dark areas, because of their contrast with the brighter areas considered above.

The dark areas cover about one-fourth of the planet, mostly in the southern hemisphere. The degree of darkness is not the same for all the dark areas, and even for a given area it varies during the course of a Martian year. The albedo is also variable, but the average value for visible light is ap-

proximately 0.16, as stated earlier. Thus, the dark regions on Mars have a little more than half the brightness of the brighter areas.

The color of the dark areas has been the subject of much discussion. On the basis of early visual observations, these areas have been described as being gray (or blue gray) in the local winter, changing to green (or blue green) in the spring and summer. There is a possibility that these reported colors may be attributed to psychophysiological effects. When a neutral-colored area is adjacent to one that has a definite color, the former often appears to have the complementary color. This could cause regions adjoining the orange and reddish areas on Mars to appear greenish or bluish. The change in apparent color of the dark areas with the season is attributed to changes in the albedo.

During the opposition of 1956, G. P. Kuiper made a careful study of Mars, including the colors of the dark regions, with the 82-inch telescope of the McDonald Observatory, Texas. He concluded that in the spring the dark areas had a neutral-gray color, with a touch of moss green in some of the equatorial regions. The greenish or blue-green colors reported by other observers, he thought, were caused by poor seeing conditions.

In order to avoid possible subjective effects, C. F. Capen, of the Jet Propulsion Laboratory, and his associates, used color filters in their observations of Mars during the apparition of 1964–65. In an extensive account of this work, Capen says: "The southern maria [i.e., dark areas] were medium to dark contrasts of dark purple and brown. The Syrtis Major was changing from a blue-green to a green-blue hue. The Trivium Charontis showed only a moderate-contrast brown coloration."

In another section, Capen reports: "The Mare Acidalium changed from its winter shades of variegated gray and brown to its spring coloration of dark gray and blue-gray shades with gray-green oases. . . . In . . . late spring and early summer, the Acidalium . . . became a very dark gray general shade with a black-green central area and large dark gray-green oases." There is obviously no simple answer to the question: What is the color of the dark areas on Mars?

Another approach to the color of these dark areas can be made on the basis of the variations of the surface reflectivity with the wavelength of the light. The quantity called the reflectivity here is a measure of the sunlight reflected back to Earth by the Martian surface when the directions of both the incident sunlight and the reflected light are almost at right angles to the surface. The results, derived from the data of A. Dollfus and J. H. Focas, are shown in figure 6.7 for both bright and dark areas of Mars. At the time the measurements were made, the dark areas were not at their darkest.

The marked increase in the reflectivity of the bright areas with increasing wavelength of the light, that is, in going from blue to red, accounts for the orange and reddish colors of

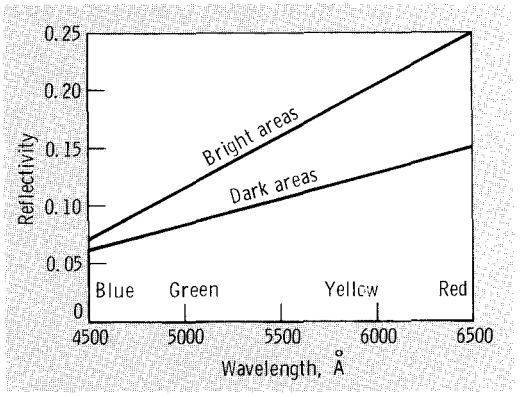


FIGURE 6.7. Reflectivity of bright and dark areas of Mars at different wavelengths. (After A. Dollfus and J. H. Focas.)

these regions. But the dark areas also exhibit an increase in reflectivity, although a smaller one, with wavelength. Hence the dominant color of these areas must be toward the red end of the spectrum. Thus, as J. B. Pollack and C. Sagan have stated: "The dark areas are red, although not as red as the bright areas." The neutral gray appearance can be accounted for, as will be seen in due course, but it is difficult to understand the blues and greens observed with color filters.

Although the bright areas differ somewhat from one another in color, each region seems to have a uniform appearance. But this is not true for the dark areas. In 1909, E. M. Antoniadi had reported, and later confirmed, what he called the "leopard-skin" structure of these areas; they consisted of darker spots on a somewhat lighter background. According to the description given by J. H. Focas in 1962, the dark areas on Mars consist of "a dusty background of a rather granular appearance patched by dark blocks, groups of spots, or isolated spots of various sizes, constituting the fundamental nuclei of dark matter in these areas." Part of a map of Mars prepared by Focas, reproduced in figure 6.8, indicates how he saw the dark areas.

Changes in the Dark Areas

An important characteristic of the dark areas on Mars is the changes they display in the degree of darkness and often in their size. Three general types of changes in appearance have been described. There are, first, widespread seasonal (or periodic) changes which occur during the course of each Martian year. Then there are localized seasonal changes which may vary from year to year. Finally, there are irregular secular changes which may persist for several years.

In the 1880's, G. V. Schiaparelli had noticed that some dark areas did not appear equally dark at successive oppositions, and he suggested that this might be a seasonal effect. Similar changes were reported by the French astronomer E. L. Trouvelot in 1884. Then in

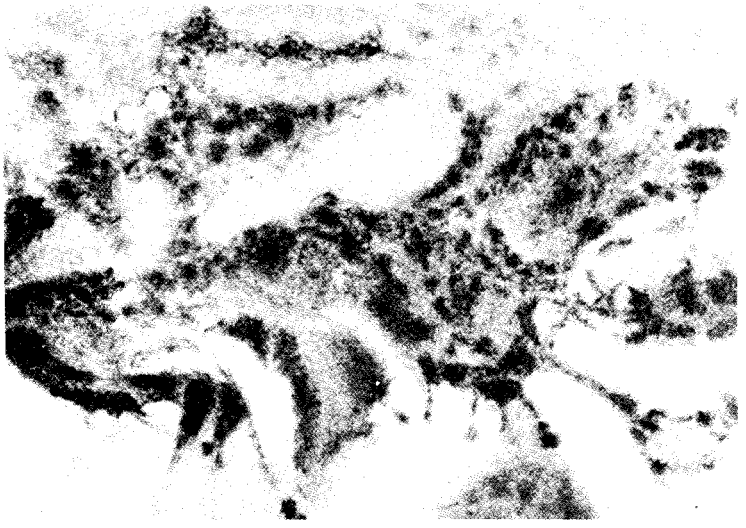


FIGURE 6.8. Granular appearance of dark areas of Mars. (Drawing by J. H. Focas; courtesy University of Chicago Press.)

1894, P. Lowell stated that "during the summer of the Martian southern hemisphere, a wave of seasonal change swept down from the poles over the face of the planet." A more complete description of the change in the dark areas was given by Lowell in 1896. "The first marked sign of change," he wrote, "was the reappearance of Hesperia [240° W, 20° S], whereas in June [1895] it had been practically nonexistent, by August it had become perfectly visible . . . but it returned in October to a midposition of visibility."

The period June through October 1895 corresponded to late spring and early summer in the southern hemisphere of Mars. The area known as Hesperia, which had been very faint in the local winter, became much darker during the late spring. But in the early summer the intensity had decreased again. This temporary darkening effect was observed by W. H. Pickering on other dark areas of Mars, and in 1924, E. M. Antoniadi reported its general nature. Just before the local spring equinox each year, the dark areas nearest the pole become darker. The darkening spreads toward lower latitudes, later and later in the spring, but the already darkened areas nearer the poles become somewhat lighter again, as they were in the winter (fig. 6.9).

An attempt at a quantitative study of this wave of darkening, as it is now called, was made by G. de Vaucouleurs in 1939. He recorded visual estimates of the changes in intensity during the spring in the southern hemisphere of Mars of various dark areas, ranging from Depressio Hellespontia, at a latitude of approximately 60° S, at the edge of the south polar cap, to Niliacus Lacus, at about 40° N, in the northern hemisphere. "The darkening," said de Vaucouleurs, "starts near the end of the southern winter at about latitude 60° S; it then spreads and crosses the equator before mid-spring, reaching latitude 40° N before the end of the southern spring."

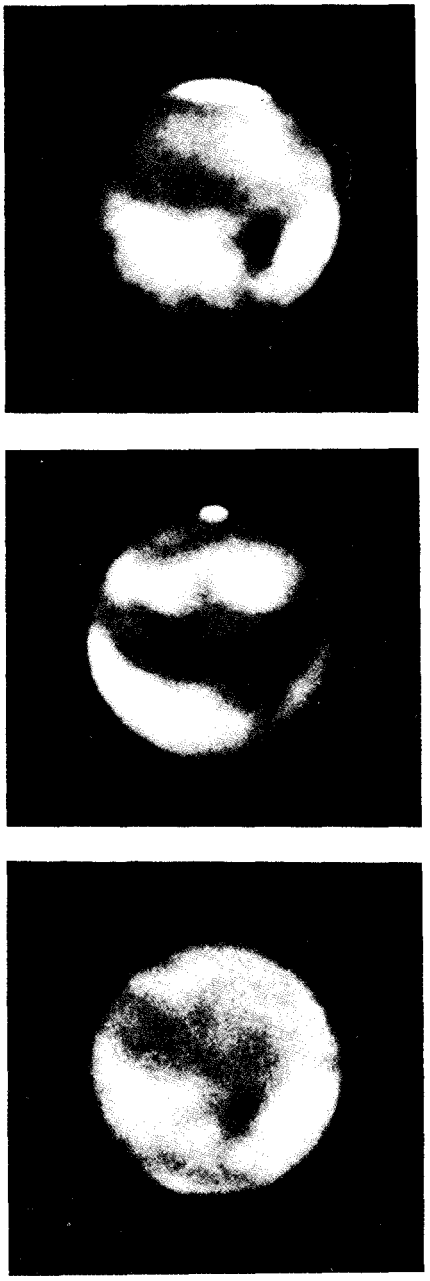


FIGURE 6.9. Photographs of the same region of Mars taken on Martian dates equivalent to April 7, July 1, and August 27. (Lowell Observatory photograph.)

In 1962, J. H. Focas reported the results of observations on the seasonal darkening made with instruments (photometers), during several Martian apparitions, at the Pic du Midi Observatory in France. A general indication of the conclusions, to which there are some exceptions, is represented in figure 6.10. The dashed arrows for each hemisphere indicate the approximate seasonal time limits of the darkening period at various latitudes, whereas the continuous arrows show when the maximum darkening is attained. The phenomena are less clearly evident in the northern than in the southern hemisphere, but they are probably similar in both hemispheres.

The wave of darkening is seen to start from each polar cap in turn, at intervals of half a Martian year, toward the end of the local winter and to spread into the other hemisphere by late spring. According to Focas, "a darkening wave can be traced . . . traveling at a rate of 30 kilometers [19 miles] per day from the circumpolar area toward the equator and extending across the equator to latitudes 22° [approximately] in the opposite hemisphere."

An interesting point which is apparent from figure 6.10 is that in the vicinity of the equator, the darkening is essentially continuous. The wave from one pole commences just about the time that the wave from the other pole is ending. The equatorial regions thus remain fairly dark essentially the whole year around. "Temperate and circumpolar areas," says Focas, "may attain a maximum intensity equal to that of the equatorial areas, but they weaken or vanish as soon as the wave has passed."

A statistical study, published in 1967 by J. B. Pollack, E. H. Greenberg, and C. Sagan, shows that, although there is a significant correlation between the latitude and the time of maximum darkening, there are nevertheless some exceptions. An outstanding example is

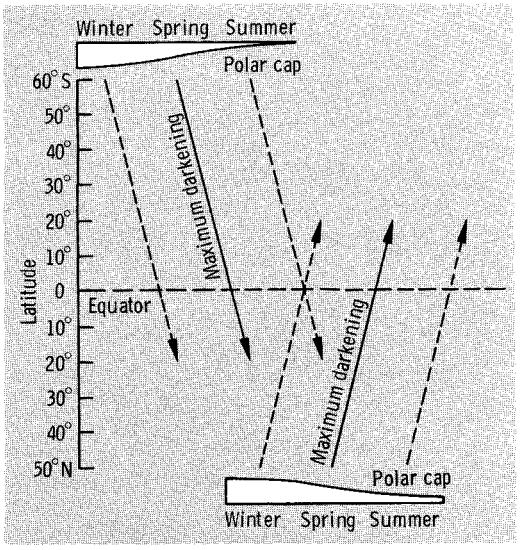


FIGURE 6.10. Wave of darkening on Mars. (After J. H. Focas.)

provided by Tithonius Lacus, which is close to the equator (5° S latitude). When the darkening spreads down from the south pole of Mars, this area begins to darken "earlier than any other dark area and reaches its maximum darkening at a time similar to those for the highest latitudes [i.e., nearest the poles]." There are either local surface variations or atmospheric conditions (or both) in the Tithonius Lacus area that are responsible for the exceptionally early darkening.

Before discussing the possible cause of the wave of darkening, the two other types of changes in the dark areas will be described because all these changes may well be related. In addition to the regular seasonal variations, there are often erratic or irregular local changes associated with the regular ones. For example, certain dark areas appear different each spring and the boundaries vary from one Martian year to another and so also do the structural details. An outstanding illustration is provided by the region called Solis

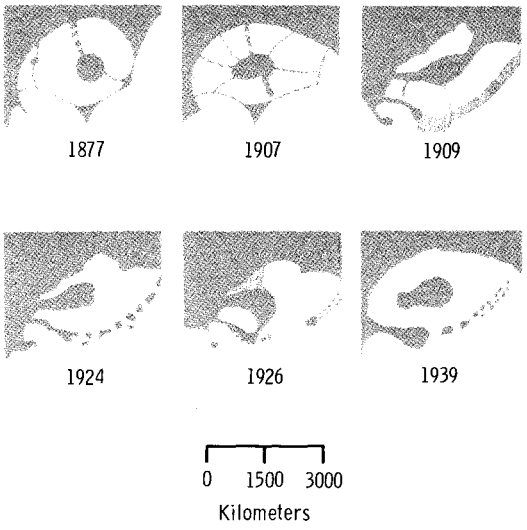


FIGURE 6.11. Appearance of the Solis Lacus region near six Martian oppositions. (After E. M. Antoniadi, "La Planète Mars", and V. V. Sharonov, "Mars.")

Lacus (90° W, 28° S). Its appearance near six Martian oppositions from 1877 through 1939 is shown in figure 6.11. The changes are evidently not permanent, and E. C. Slipher reported that in 1958 and 1960 the Solis Lacus area closely resembled its appearance in 1907.

Among other regions showing irregular behavior are the Nepenthes-Thoth area, Utopia, Trivium Charontis, and Nilokeras. In 1909, Nepenthes-Thoth was almost invisible, but by 1939 it had developed into a conspicuous dark band. Although it was described as a slight dusty streak in 1952, at the next apparition in 1954 it had become a large dark area. In later years Nepenthes-Thoth was much less marked again. According to A. Dollfus, "it seems that all Martian [dark] features are, sooner or later, subject to such changes."

The third type of change observed in the dark areas is represented by the long-term (or secular) changes. These variations often arise quite suddenly, that is, between two successive oppositions, and may last for several years. In 1941, A. Dollfus observed a broad and dark area in the Boreosyrtis (290° W, 55° N) and Dioscuria (320° W, 50° N) regions which had not been recorded on any earlier maps. It has been observed at subsequent Martian apparitions, although its intensity has varied from one opposition to the next. Another example is provided by the dark Mare Cimmerium which extended its northern edge by almost 1000 kilometers (620 miles) in 1939, and has remained so since. The southern boundary, along the bright area Zephyria, changed at the same time, but it returned to normal in 1948.

Occasionally, a large new dark area of considerable size develops within a bright region. E. C. Slipher reported in 1954 that a previously unobserved dark area of nearly 2.5 million square kilometers (600 000 square miles) had developed east of Thoth and remained essentially unaltered in later years (fig. 6.12). Many other cases of secular changes of the dark areas on Mars have been

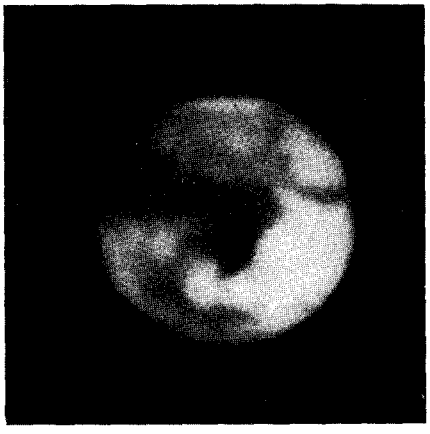


FIGURE 6.12. Large dark area (bottom left) first observed in 1954. (Lowell Observatory photograph.)

reported, but the illustrations given above are sufficient to indicate the nature of the phenomenon.

In spite of the changes in the Martian surface, both seasonal and secular, the main features of the planet are permanent in nature. It is true that the same details cannot always be seen but, in general, the prominent named areas appear in the same locations from one apparition to another. Were this not so, it would not be possible to prepare maps of Mars, such as the one in figure 2.21.

Some regions have remained the same, apart from seasonal darkening, for over 100 years. An outstanding example is Syrtis Major which Huygens represented in his crude drawing of 1659. Another is the dark Sinus Sabaeus which stretches across the planet for some 3000 kilometers (1860 miles) almost parallel to and south of the equator. It is clearly seen in the map of W. Beer and J. H. von Mädler of 1840 (fig. 2.15) and in R. A. Proctor's map of 1867, as well as in modern maps of Mars. There are sometimes minor differences in size and degree of darkness, but the reason is usually that the seasons are not always the same on Mars at successive oppositions.

Explanation of the Wave of Darkening

Interpretation of the changes in the dark areas of Mars is one of the most intriguing and controversial problems relating to this planet. An examination of figure 6.10, where the changes in the polar caps throughout most of the year are indicated at the top and bottom, shows that the wave of darkening commences at high latitudes, nearest the pole, at about the time the polar cap begins to recede. The wave then continues to move to lower and lower latitudes, while the polar cap gets smaller and smaller in size during the local spring.

It would seem, therefore, that there may be a direct connection between the disappearance of the polar cap and the steady advance of the wave of darkening toward and across the equator. If the polar cap consists (or contains a significant amount) of water, then it might be expected that the gradual transfer of water vapor from higher to lower latitudes is responsible in some way for the darkening. Although different ideas were expressed concerning the function of the water, the general point of view was accepted almost universally until 1957.

Perhaps the most popular interpretation of the wave of darkening, because it appears to be so simple and straightforward, is that the darkening is related to the presence of a primitive form of vegetation. This possibility was first indicated by E. L. Trouvelot when he wrote in 1884 that "one could believe . . . these changing grayish areas are due to Martian vegetation undergoing seasonal changes." During the Martian winter, this vegetation would be dormant because of the low temperature and the lack of water.

As the solidified water from the polar cap started to vaporize (and possibly to liquefy) at the end of the local winter, the water vapor would be carried to lower latitudes by the prevailing winds. The arrival of water in some form, combined with the simultaneous increase in temperature as winter turned to spring, would cause the vegetation to revive. The resulting change in the character of the surface would then appear as a darkening. Later in the year, as the amount of water vapor in the atmosphere declined, the vegetation would become dormant again and the surface would appear less dark.

It was pointed out by P. Lowell in 1895 that the foregoing hypothesis would imply a situation on Mars quite different from that on Earth. On Earth, the spring revival of vegetation commences at low latitudes, near

the equator, where it is warmest, and extends to higher latitudes later and later in the year. On Mars, however, the reverse would be true. Revival would start near the pole and proceed toward the equator. The reason for this difference is that on Earth there is generally an ample supply of moisture and the revival of plant life is usually dependent on the temperature. But on Mars, water is the limiting factor, and this would become available first in regions nearest the pole where the main supply is concentrated.

What was at one time regarded as a cogent argument in favor of the vegetation hypothesis was put forward in 1950 by the Estonian-born astronomer E. J. Öpik, then in Northern Ireland. Mars is subject to occasional severe duststorms which may cover extensive areas of the planet and last for several weeks (ch. VII). Öpik pointed out that unless there was some means of regeneration, the dark areas would have become completely covered by the yellowish-orange dust during the course of time. Dark regions do appear lighter after a duststorm, but they recover their original appearance within a few weeks. It seemed that the growth of a form of plant life would be the most obvious way to explain this recovery. There is, however, an alternative possibility which will be examined shortly.

The idea that vegetation is responsible for the seasonal wave of darkening can also be applied to account for the other types of changes observed in the dark areas. The fading away of an existing dark region can be explained by the rapid extermination of the vegetation. It is much more difficult to understand how a new dark area, sometimes covering several hundred thousand square miles, could form within a single Martian year.

In contrast to the vegetation (or organic) hypothesis, some inorganic theories have been proposed to account for the role played by water in the darkening phenomenon. In 1912, S. Arrhenius suggested that the water vapor interacts with substances present in the Martian soil, and that this is responsible for the darkening. Somewhat similar effects are known to occur in certain terrestrial deserts, but the quantity of moisture required is probably much more than would be available in the Martian atmosphere. The views of Arrhenius have received some support, but on the whole they have not been found to be satisfactory.

Another inorganic mechanism, also involving water to account for the wave of darkening, was proposed by J. Otterman and F. E. Bronner of the General Electric Co.'s Missile and Space Division in 1966. "It is suggested that freezing [of moisture in the soil] during the Martian afternoon and evening produces one or more types of surface microrelief features." Such features, referred to as microhills, have been observed to form on Earth as a result of the freezing of moist soils. By increasing the complexity, that is, the porosity and roughness, of the surface, the microhills cause the reflectivity to diminish and the optical darkening to increase.

During the daytime, as the surface warms up, the frozen water would be vaporized and gradually transported to lower latitudes where it would be used again to produce darkening. The microhills would retain their structure for some time, even when dry, but would be eventually destroyed by wind erosion, and the surface would return to its original state and lighter color.

The theory outlined above requires that water be available in liquid form, at least for a short time each day. At the low partial pressure in the Martian atmosphere, the water would normally deposit as solid hoarfrost on cooling. Such deposition would occur in the evening or at night. The next day, when

warmed by the Sun, part of the hoarfrost would vaporize by sublimation, but part might turn into liquid water and be absorbed and held by capillaries in the soil. Upon freezing, the microhills would then be produced.

The great drawback to the microhill hypothesis, and in fact to any explanation of the wave of darkening involving water in some way, is that the reported average content of water vapor in the Martian atmosphere, about 14 microns of precipitable water, is so low. According to J. Otterman (1967), the minimum quantity of precipitable water required for the microhill theory is 200 microns. He suggested that spectroscopic observations be made at various locations on Mars to see if the quantity of water in the atmosphere is indeed as low as has been reported. If subsequent investigations establish that the amount of precipitable water is everywhere substantially less than 200 microns, the microhill explanation of the wave of darkening would have to be abandoned.

Another difficulty, which is independent of the question of the water content of the Martian atmosphere, has been pointed out by C. Sagan. He has called attention to the fact that the microhills would be larger than the particle sizes derived from photometric and polarimetric observations.

Because the amount of water vapor in the Martian atmosphere appears to be quite small, theories of darkening which are independent of the presence of water are now attracting interest. The basis of these theories is that light-colored dust from the bright regions is responsible for the changes, both seasonal and secular, of the dark area of Mars. This idea was suggested in 1957 by G. P. Kuiper, who thought that the dark areas might be lava fields covered with dust. Removal of the dust by the wind, such as commonly occurs on terrestrial lava flows, would then result in a darkening of the surface. "The Martian maria [i.e., dark areas] could change their visibility with the seasons depending on the atmospheric circulation."

A similar view was expressed independently by V. V. Sharonov of the U.S.S.R. in 1958. "The air currents in the [Martian] atmosphere," he wrote, "vary from season to season, depositing dust at some time of the year and blowing it away at other times. Thus, for instance, the inherently dark surface . . . may brighten at a definite time of the year as a result of settling of light-colored dust blown over from the desert areas."

The interpretation of the changes in appearance of the dark areas on Mars in terms of particle size and winds has been developed by D. G. Rea in 1964 and by J. B. Pollack and C. Sagan in subsequent years. Only the smaller, light-colored particles would be carried by the winds; the larger particles, on the dark areas, would merely saltate (bounce) along the surface. The marked changes in appearance, therefore, would be expected only on the dark regions. There is, however, some evidence that the same bright areas brighten while the dark areas are darkening. Both of these effects would be caused by the wind blowing the finer and lighter colored particles from the dark to the bright regions.

"Photometry and, particularly, . . . polarimetry [studies]," say Pollack and Sagan, "indicate that the principal event of the wave of darkening is a change in the mean particle size, . . . with no substantial change in composition."

One question may well be asked in connection with the theory of windblown dust. Is the regular nature of the wave of darkening, from one year to another, consistent with the behavior expected from a wind system? It would probably not be on Earth, but it may be so on Mars where the meteorological structure is less complex.

If the removal of dust from the dark areas

and its subsequent replacement by the wind is responsible for the wave of darkening, then the correspondence in time of this effect with the recession of the polar caps is purely coincidental. Both phenomena, although independent of each other, might nevertheless be caused by seasonal temperature changes in the atmosphere and of the surface. Local wind conditions could perhaps explain why Tithonius Lacus, near the equator, starts to darken before areas that are much closer to the south pole. Certainly the darkening mechanisms involving water are inadequate in this situation.

The irregular seasonal variations in the dark areas would seem to find a ready interpretation in the hypothesis that darkening is a result of the removal of light-colored dust. Irregularities in the surface, local differences in elevation, and associated variations in the wind pattern could produce a different appearance from year to year. The formation of large, new dark areas, in the secular changes, could result from the removal of a thin layer of dust by the wind. That such changes take place within a single Martian year would, then, not be surprising.

Finally, when a dark area is covered with yellow dust as the result of a Martian dust-storm, removal of the dust by the wind would account for the relatively rapid restoration of such an area to its original dark appearance. The removal would be facilitated if the dark areas, as suggested later, are elevated regions with gentle slopes. Thus, what was regarded as one of the strongest arguments in favor of the vegetation theory of darkening appears to lose much of its force.

Nature of the Dark Areas

There seems little doubt that the surface of the bright areas of Mars consists to a significant extent of some form of hydrated ferric oxide, with probably some silicate minerals. The situation with regard to the dark areas, however, is less definite. For many years there was not much speculation concerning the composition of the darker surface material, but since about 1964 there has been growing evidence for the view that essentially the same substances are present on the surfaces of both bright and dark areas. The difference in appearance is ascribed to the particles being larger on the dark surfaces. Such particles would reflect less light than the smaller particles on the bright areas and so they would appear darker. It is well known that many materials which are dark red, almost black, in color, because they have very poor reflectivity, appear orange and yellow in powdered form.

It will be recalled from figure 6.7 that the reflectivities of both bright and dark areas on Mars increase as the wavelength of the light increases from blue to red. The smaller increase exhibited by the dark regions is just what would be expected if they consisted of larger particles of the same material as covered the bright areas. The larger particles will absorb more, and reflect less, of the sunlight at the red end of the spectrum than will the smaller particles in the bright areas. Because they reflect about the same in the blue region, the larger particles will tend to give the surface a bluish-gray appearance.

It is of interest, too, that polarization measurements made by A. Dollfus show relatively little difference between light and dark areas. The dots in figure 6.13 represent the differences between the polarization of dark areas at different times of the year and the average for bright areas measured at a phase angle of 25°. It will be noted that, although there appear to be definite seasonal changes in the polarizations of the dark areas, the values differ by not more than ± 0.002 from the polarization of a bright area. In fact, for most of the local spring and summer, the

polarization of the dark regions at a phase angle of 25° is essentially identical with that of the bright areas.

In a detailed review of the photometric (reflection) and polarization properties of the surface of Mars completed in 1967, J. B. Pollack and C. Sagan have concluded:

> [Hydrated] ferric oxides are a major constituent of the bright areas . . . [and] also . . . of the dark areas. [This] . . . follows from the near identity of the refractive indices of bright and dark areas [as calculated from reflectivity data. The] . . . agreement . . . [arises] from the very low contrast in the blue, violet, and ultraviolet, and from the fact that the polarization . . . of the dark areas can be derived from that of the bright areas merely by increasing the mean particle size. Even during the seasonal darkening of the dark areas, their index of refraction remains almost the same as for the bright areas.

It should be emphasized that the idea developed above, that the bright and dark areas of Mars have the same composition but differ in particle size, is only a theory, although admittedly an attractive one. It is thus necessary, as with so many other aspects of the planet Mars, to keep an open mind on this subject.

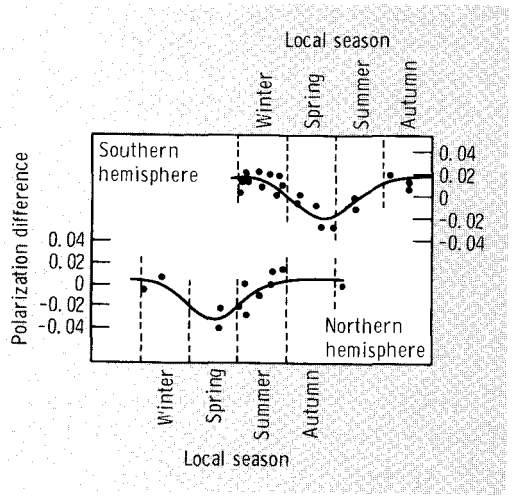


FIGURE 6.13. Seasonal variation in polarization difference between light and dark areas. (After A. Dollfus.)

Elevations of Bright and Dark Areas

On the Moon, the darker regions are undoubtedly at a lower elevation than the bright, highly cratered areas. For about a hundred years, the general opinion among astronomers had been that an analogous situation existed on Mars. The dark areas were considered to be low-lying regions, whereas the bright ones were thought to be uplands. This view was based to a great extent on observations of the formation of clouds, the deposition of frost, and the disappearance of the polar caps. The word "frost" is used here in a general sense to include both water frost (hoarfrost) and solid carbon dioxide.

In the first place, it is reported that white clouds on Mars seem to be formed preferentially, and seem to remain, over certain bright areas (ch. VII). Furthermore, frost deposits, which are seen to be present on the western limb of the planet as it emerges from night into daylight, tend to be localized in the bright regions. Finally, the south polar cap is often extended locally into such bright areas as Hellas and Argyre. In addition, the remains of this polar cap are located over a bright region, whereas the south pole itself, from which the cap disappears in the summer, is dark. The bright areas are assumed to be colder, at a higher altitude than the dark regions.

If the frost deposits and the polar caps consist mainly of solid carbon dioxide, then at least some of the foregoing arguments may be invalid. (They may also be invalid even if the polar caps consist of condensed water, such as hoarfrost, and not carbon dioxide.) It is probable that the Martian atmosphere is mostly carbon dioxide gas, so the tendency for carbon dioxide frost to deposit on the ground

or to cause a solid carbon dioxide polar cap to form would be determined by the atmospheric (barometric) pressure, rather than by the temperature. In these circumstances, the bright areas would be those where the atmospheric pressure is highest, and they would be at low rather than at high altitude.

Regardless of the nature of the Martian frost and polar caps, it is by no means certain that the higher elevations on Mars will be significantly colder than adjacent low-lying areas. There are three main reasons why this situation exists on Earth. First, the "greenhouse" effect, whereby significant quantities of heat radiation from the ground are trapped by the atmosphere, decreases with increasing altitude. Second, because of the up-and-down slopes in a mountainous region, the radiation from the Sun is spread more thinly over a larger area than on flat, low-lying ground. Third, the air is cooled as a result of expansion when it rises to higher elevations where the atmospheric pressure is lower.

C. Sagan and J. B. Pollack (1966) have argued, however, that none of the circumstances described above as existing on Earth is of great significance on Mars. On the whole, it appeared that the highlands on the latter planet might be no more than a few degrees cooler than the nearby lowlands. The common belief that the higher elevations on Mars must be substantially colder than low-lying areas might be quite unjustifiable, and conclusions based on this belief may consequently be incorrect.

The foregoing considerations indicate that the bright areas on Mars are not necessarily uplands. Arguments will now be examined which suggest that they are actually at lower elevations than the dark areas. In the first place, D. G. Rea pointed out in 1964 that a variation in size of the surface particles combined with a range of elevations could result in a fractionation (or separation) by the winds over a period of time. If this occurred, there might be a tendency for the smaller particles to collect on large, flat, low-lying areas, and these would be bright in appearance. On high, flat regions and gentle slopes, the particle-size distribution would depend on the atmospheric circulation, but there would be a preference in favor of the larger particles. The elevated areas would then be darker in color. Seasonal exchange of the smaller particles by the wind, between the higher (bright) and lower (dark) areas, would account for the observed changes in the dark regions.

Another argument was put forward in 1965 by R. A. Wells, University of London, based on the observations reported by A. Dollfus and J. H. Focas that certain isolated white (condensation) clouds in the Martian atmosphere tend to form and remain stationary over bright areas. Such clouds seem to be adjacent to, and alined with, the boundary between a bright area and a dark region. They occur, for example, at the boundaries of the dark Sabaeus Sinus with the bright Deucalionis Regio (to the south) and Edom (to the north).

On Earth, similar stationary clouds are known to form on the lee side of a high ridge, the side protected from the wind. It is by no means certain that the conditions for cloud formation on Mars are the same as those on Earth; but if they are similar, then the bright areas over which the clouds are observed are at a lower elevation than the adjoining dark areas.

C. Sagan and J. B. Pollack have presented other meteorological evidence that appears to favor the idea that the dark areas on Mars are elevated regions. It will be seen in the next chapter that the yellow dust clouds, which form occasionally in the Martian atmosphere, develop over the bright areas and then are

FIGURE 6.14. Apparent deflection of dust clouds by dark areas. (Drawings by A. Dollfus, Annales d'Astrophysique, Vol. 28, p. 722 (1965).)

carried, presumably by winds, across the planet. There are indications that, although the yellow clouds sometimes obscure dark areas adjacent to the bright region where they originated, the darkest of such areas are rarely crossed by these clouds. In fact, the dust clouds often appear to be deflected by dark areas (fig. 6.14). The major dark areas thus seem to constrain the possible paths of a dust-storm. It is to be expected that the dust particles would be carried by the winds mainly along valleys and across lowlands, in general. If this is the case, then the dark areas, which the clouds seem to avoid, are uplands.

During the oppositions of 1963 and 1965, R. M. Goldstein, using the facilities of the Jet Propulsion Laboratory at Goldstone, Calif., studied radar signals at a wavelength of 12.5 centimeters, which were reflected from Mars as it rotated. In this manner, the radar tracked a swath around the planet between latitudes of approximately 10° N and 30° N, cutting across several prominent dark areas on the surface as well as many bright regions.

Region	Longitude	Latitude	Average slope, deg	Maximum elevation	
				Kilometers	Feet
Moeris Lacus	270° W	8° N	3	16. 5	51 000
Niliacus Lacus	30° W	30° N	3	15	49 000
Syrtis Major	290° W	10° N	4	12	39 000
Nodus Lacoöntis	250° W	20° N	2	7	23 000
Trivium Charontis	198° W	20° N	2	5	16 500
Nepenthes	260° W	20° N	1	4	13 000

The results were reported and interpreted in 1966 in conjunction with C. Sagan and J. B. Pollack.

It was observed that high radar reflection seemed to be associated with the dark areas on Mars, although the positions of maximum radar reflection were always displaced somewhat from the locations of the dark areas. An examination of the factors which might cause such enhanced reflection, taken in conjunction with other considerations, led to the conclusion that the dark regions are highlands. The mean slopes and maximum elevations above adjacent bright areas of a few dark regions on Mars, as estimated from the radar reflections, are given in the table on page 123. It should be noted that the slopes and elevations quoted are uncertain by a factor of about 2. Nevertheless, the results imply that the major dark areas have mean slopes of a few degrees and their maximum elevations above the nearby bright areas may be as much as 10 kilometers (33 000 feet) or more.

The dark areas, such as Nodus Lacoöntis, Trivium Charontis, and Nepenthes, which are known to exhibit prominent changes in appearance, are seen to have gentle slopes, between 1 and 2 degrees, and moderate heights of about 5 kilometers (16 500 feet). Moreover, these relatively small dark regions are completely (or largely) surrounded by bright areas. The three existing conditions—gentle slope, moderate elevation, and adjacent bright regions—are just those which may be expected to favor coverage by and removal of small particles carried by the wind.

The Moeris Lacus, Niliacus Lacus, and Syrtis Major regions, which do not undergo significant secular changes, on the other hand, have steeper slopes, higher elevations, and are only partially bordered by bright areas. The conclusions from the radar studies are thus in harmony with the view that secular changes in dark (elevated) areas are the result of particles blown off (or onto) them from adjacent bright (lower) regions.

The largest maximum altitudes given above for Mars, about 15 or 16 kilometers (50 000 feet), may be taken as the difference in elevation between the lowest and highest surface areas. On Earth, this would be equivalent to the vertical distance between the ocean bottom and the top of a high mountain. It is of interest that this distance has much the same magnitude as the difference between the lowest and highest elevations on Mars.

From radar studies made in 1967, G. H. Pettingill, of the Massachusetts Institute of Technology, arrived at altitude differences and slopes on the Martian surface of the same order of magnitude as those given above. But the correlation between dark areas and high elevations was much less clear. The measurement technique used by Pettingill differed from that employed by Goldstein, however, and sources of possible discrepancy between the results given by the two methods are being investigated.

An observation which may have significance concerning the relationship between elevation and the bright and dark areas of Mars was reported in 1968 by P. D. Lowman of the Goddard Space Flight Center. He pointed out that in the photographs of Earth taken from the manned Gemini spacecraft, the lower and level desert areas appear brighter than the mountainous regions. Although a comparison of photographs of Earth taken at a distance of the order of 100 miles with the appearance of Mars in the telescope may not be completely justifiable, the results do suggest the possibility that the bright and dark areas on Mars may be at lower and higher elevations, respectively.

A few words may be said here about the possibility of determining elevations on Mars by methods other than that based on radar

reflection. Differences of altitude have been measured on the Moon by observing shadows cast near the terminator at times, close to the quarter phases, when the Sun is low on the lunar horizon. As a result of applying the same principle to Mars, P. Lowell concluded, in 1906, from his failure to observe shadows at the terminator, that abrupt heights on the planet did not exceed some 0.8 kilometer (2500 feet).

A reexamination, reported in 1961, of Lowell's estimates was made at the New Mexico State University by C. W. Tombaugh, the discoverer of the planet Pluto. He concluded that the smallest vertical height on Mars that could be detected from the shadow at the terminator is about 8.4 kilometers (27 500 feet). "This means of mapping relief [on Mars]," wrote Tombaugh, "is beyond ground-based capabilities." Moreover, the calculation of elevation from shadow lengths assumes steep slopes. Elevated areas on Mars with gentle slopes, such as appear to be indicated by the radar measurements, would not be detected by the shadow technique even if considerably higher than 8.4 kilometers.

THE CANALS AND CRATERS

The Canals of Mars

The canals have been the subject of more heated and prolonged discussion than any other aspect of the planet Mars. Many reputable astronomers, following G. V. Schiaparelli, claim to have seen the long and narrow linear markings on the surface, whereas others, equally reputable, have been unable to confirm these observations. There are some differences in the views of those who affirm that the canals exist, indicating that there is a degree of subjectivity in their identification. Thus, in 1906, W. H. Pickering wrote: "Some observers like [A. E.] Douglass draw them [the canals] very wide; others, like [P.] Lowell . . . extremely narrow. The writer [W. H. Pickering] gives them an intermediate breadth . . . The long canals never appear to him to be as narrow as they are drawn by Lowell."

The failure of many astronomers to detect the Martian canals was attributed by P. Lowell to poor seeing conditions. He thought that only on relatively rare occasions, when the atmosphere is clear and free from turbulence, can the canals be observed. "When a fairly acute eyed observer sets himself to scan the telescopic disk of the planet in steady air," said Lowell, ". . . he will . . . of a sudden become aware of a vision of a thread stretched across the orange areas. Gone as quickly as it came, he will instinctively doubt his own eyesight . . . [then] with the same startling abruptness, the thing stands before his eyes again."

The opposite point of view, that the apparently continuous linear features are the result of poor resolution, was put forward by E. M. Antoniadi in 1929. To illustrate this point, he presented the drawings reproduced in figure 6.15 of the Elysium region of Mars. The one at the left is by G. V. Schiaparelli based on observations made with telescopes of 21.8- and 49-centimeter (8.5- and 19-inch) aperture between 1877 and 1890, whereas the one on the right was made by Antoniadi, using an 83-centimeter (33-inch) instrument between 1909 and 1926.

In the same connection, two drawings made by A. Dollfus, with a 60-centimeter (24-inch) refracting telescope, of the Syrtis Major region in 1948 are of interest. The left-hand drawing in figure 6.16, made under conditions of average seeing, appears to show the double canals Thoth and Nilosyrtis. With perfect seeing, as in the drawing on the right, the continuous lines (and dark areas, in general) are broken up into light and dark

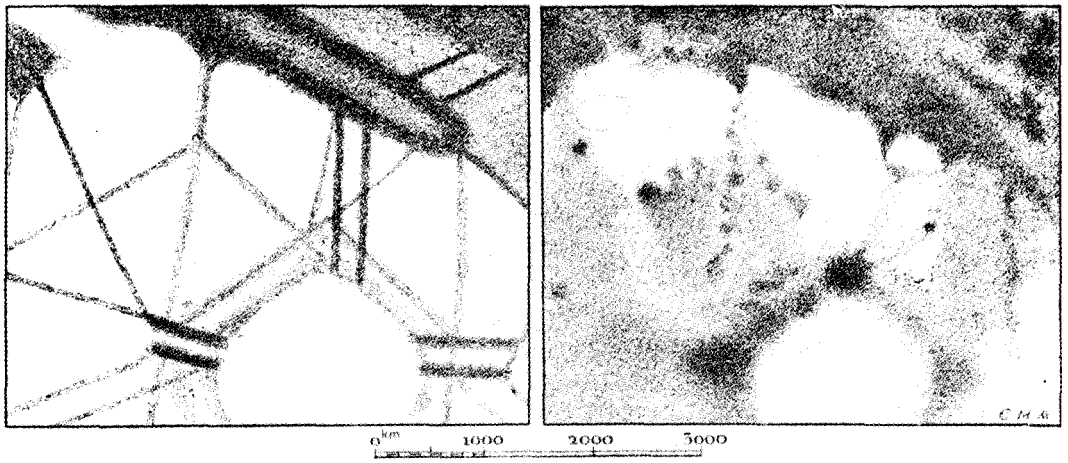


FIGURE 6.15. Comparison of drawings made by Schiaparelli (left) and Antoniadi (right). (From E. M. Antoniadi, "La Planète Mars.")

patches. In the words of Dollfus: "Canals may sometimes be noted across ocher-colored deserts; they appear only when the seeing is not very good, either as floating filaments or as soft bands. When the seeing becomes excellent, one sees at the same place only small spots or markings, more or less aligned."

At the present time, the majority of astronomers would probably agree that the extensive geometric network of canals drawn by G. V. Schiaparelli and by P. Lowell (ch. II) does not exist. Nevertheless, there definitely appear to be a number of linear features on the surface of Mars. They are probably at least 100 kilometers (62 miles) in width and, in detail, largely discontinuous and irregular. The seasonal variations in the appearance of the canals are quite similar to the changes

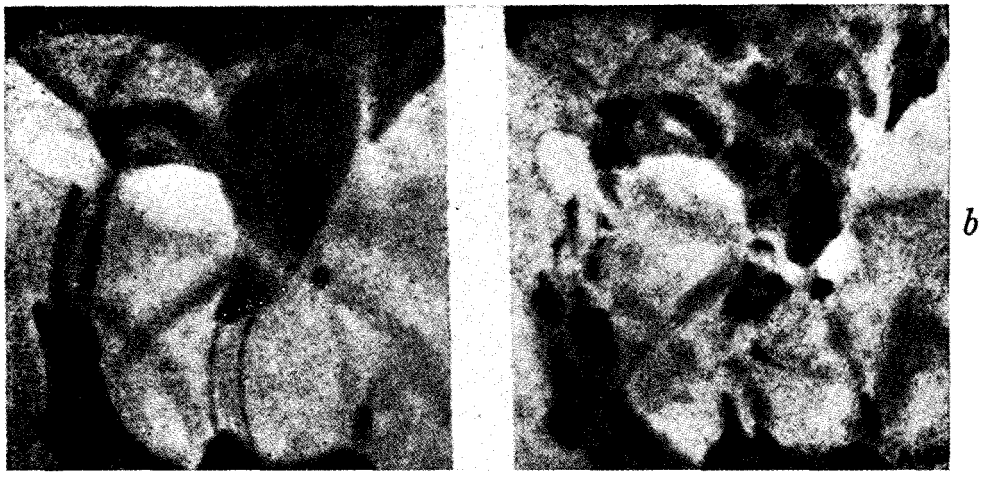


FIGURE 6.16. Drawings made by Dollfus under conditions of average (left) and perfect (right) seeing. (Courtesy University of Chicago Press.)

exhibited by the large dark areas during the wave of darkening.

Thus, Schiaparelli noted that in the local winter the canals are faint and often invisible, but during the spring they become darker. The darkening, as Lowell reported, begins near the poles and travels toward the equator, just as it does in the other dark regions. At some apparitions, often separated by several years, the canals are broad and distinct, whereas at other times they are almost invisible. G. de Vaucouleurs stated in 1966 that "the Nilosyrtis, the largest and darkest canal on Mars throughout the latter half of the Nineteenth Century, is now visible only with difficulty as a faint shadowy band, resolvable into small irregular spots." It seems probable, therefore, that the canals are relatively long and narrow, almost linear features, having the same nature as the dark areas, that run across the bright regions of Mars.

Whether the canals are depressions or elevations presumably depends on whether the Martian dark areas are at low or high altitudes, respectively. As already explained, both points of view have been propounded. According to C. W. Tombaugh (1959), for example, the canals may correspond to faults or fractures several miles wide in the Martian crust. The color changes at different seasons are attributed to vegetation growing in the low-lying regions.

On the other hand, C. Sagan and his associates claim that radar observations indicate that the canals rise above the bright areas which they cross, attaining the same maximum elevation as some of the larger dark areas, roughly 6 kilometers (almost 20 000 feet). The slopes of the canals, which are estimated to be greater than the usual 3 or 4 degrees, are steeper. If the theory is accepted that the color changes of the dark areas are caused by fine particles, deposited or removed by the wind, the marked secular variations shown by the canals can be readily accounted for by the proximity of bright areas.

A novel suggestion concerning the Martian canals was made in 1964 by F. A. Gifford of the U.S. Weather Bureau Research Station at Oak Ridge, Tenn. He contended that the deserts of Mars should contain long, straight, and narrow systems of sand dunes whose appearance would resemble the observed canals. Dune systems of this kind occur on Earth in the deserts of Africa and Arabia. One in Libya, for example, is reported to be about 650 kilometers (400 miles) long with an average width of roughly 8 kilometers (5 miles). On Mars, it is expected that the winds which produce the sand dunes have a much higher velocity than on Earth, because of the lower atmospheric pressure. The dune systems would thus probably be longer and wider on Mars. The changes in appearance of the canals might still be accounted for by the size of the particles on the surface.

It was hoped at one time that the pictures taken by the Mariner IV spacecraft when it was close to Mars in July 1965 would throw light on the nature of the canals. Unfortunately, this has not been the case for several reasons. In the first place, the spacecraft crossed the northern hemisphere, where most of the canals are located, during the local winter when the canals are very faint or invisible against the bright background. Second, each view taken by the Mariner IV cameras was of an area approximately 320 kilometers (200 miles) square. This is similar to the width of many canals, and so each picture is equivalent to little more than a very small area on a map and not much more than a point in a telescopic image of Mars. Finally, the position of the Sun when the Mariner IV pictures were taken was not too favorable for bringing out the contrast between different surface areas.

Nevertheless, examination of the Mariner IV pictures has revealed a number of what appear to be linear topographical features known to geologists as lineaments. These are, however, much narrower than the canals. They could not have been detected from Earth by the best telescopes even under ideal viewing conditions. There are similar lineaments on the Moon, as seen in pictures taken by the Orbiter spacecraft, and they are known to exist on Earth. It is generally considered that linear systems of this kind result from stresses in the crust produced by changes in the rotational velocity of the planet (or Moon) that have occurred over long periods of time. Whether or not the canals are related to the lineaments is not known.

Martian Craters

A striking feature of the photographs of Mars taken by the Mariner IV spacecraft in July 1965 (ch. XIII) was the evidence of a highly cratered surface, surprisingly similar to that of the Moon. Although this discovery aroused considerable interest when it was reported, the possibility of craters on Mars had been predicted several years earlier. It was mentioned, for example, in 1944 by a science writer D. L. Cyr in a book entitled "Life on Mars." Later, around 1949 and 1950, three astronomers, R. B. Baldwin, C. L. Tombaugh, and E. J. Öpik, independently, suggested that there might be craters on the surface of Mars. This view was based largely on the proximity of the planet to the asteroidal belt, so that formation of craters by impact with asteroidal and related fragments was to be expected.

Although there are some indications of craters in the second picture from Mariner IV, they are more clearly seen in picture 3 (fig. 13.12).[2] The failure to observe craters

[2] In all the Mariner IV photographs, north is at the top of the picture.

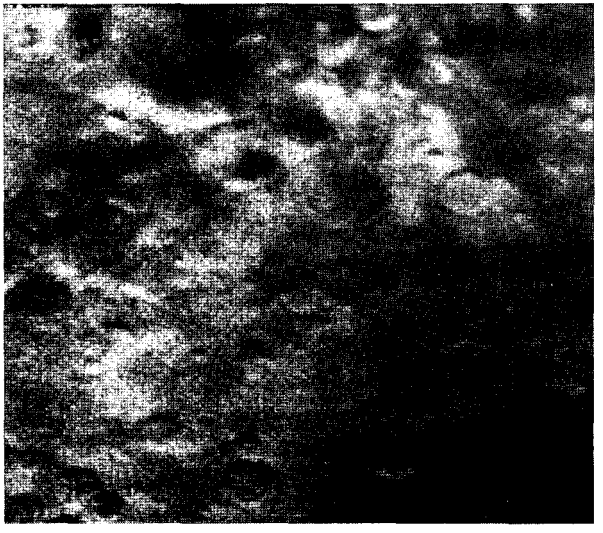


FIGURE 6.17. Mariner IV Picture No. 7; north at top.

FIGURE 6.19. Mariner IV Picture No. 11; north at top.

in some photographs is probably due to the local light and other conditions rather than to an actual absence of craters. The resemblance to the lunar surface is apparent in figure 6.17, which is Mariner IV picture 7. The area viewed is wholly within the southeastern part of the bright region Zephyria. The dimensions are about 290 by 290 kilometers (180 by 180 miles), so that the area is roughly 84 000 square kilometers (32 400 square miles). Approximately 50 possible craters have been counted in this area. Picture 9, reproduced in figure 6.18, shows a highly cratered region which is largely or wholly within the dark Mare Sirenum. As far as can be determined, there is no great difference in crater density on the light and dark areas of Mars viewed by Mariner IV.

Figure 6.19, which is picture 11, is principally within the Mare Sirenum, but the top right (northeast) corner is probably in the lighter Atlantis area between the Mare Sirenum and Mare Cimmerium. This photograph is of special interest because it shows the largest, well-defined, although not entirely complete, crater observed. Its diameter is about 175 kilometers (109 miles). There is some possible evidence of larger craters, but it is not definite enough to permit positive identification. Incidentally, a marked lineament, starting at the bottom left corner and extending at least 200 kilometers (125 miles) in an almost northeasterly direction, is an important aspect of picture 11.

Craters were detected in at least 14 of the Mariner IV photographs, but only one more, picture 14, is reproduced here (fig. 6.20). The area shown is wholly within the bright region Phaethontis. The features of special interest are the light-colored rings which are thought to be the rims (or inner walls) of craters partially covered by frost or snow, either of water or carbon dioxide. Some of the rings are incomplete in the northwest

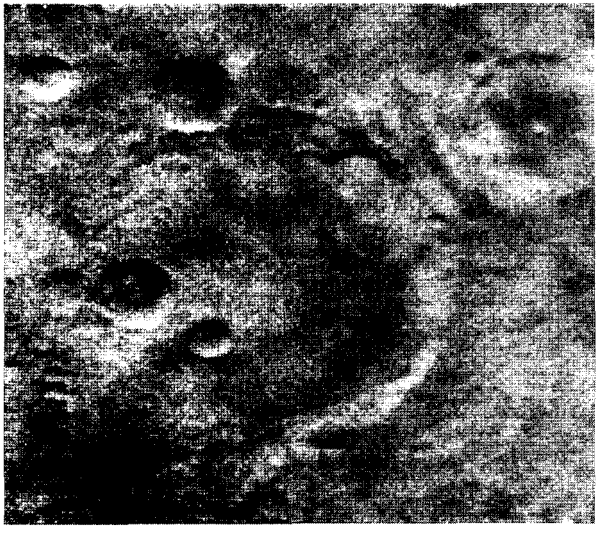


FIGURE 6.18. Mariner IV Picture No. 9; north at top.

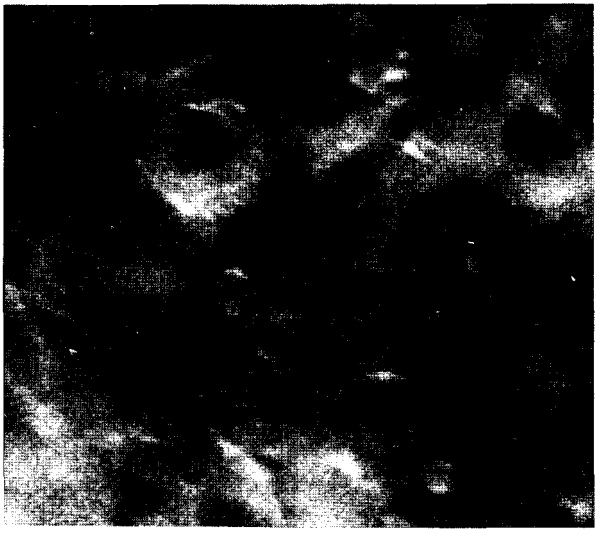


FIGURE 6.20. Mariner IV Picture No. 14; north at top.

(upper left) quadrant, as would be expected if the morning Sun had evaporated the thin deposit of frost formed during the night. It was late midwinter in the locality where the picture was taken and the area was only some 10° to 15° latitude from the edge of the south polar cap.

A few craters as small as 3 kilometers (1.9 miles) in diameter have been observed in the Mariner IV pictures, but the great majority lie within the range of 5 to 175 kilometers (3.1 to 109 miles) diameter. In pictures 7 through 12, which are the most useful from the standpoint of clarity, 206 craters with diameters greater than 5 kilometers have been identified as definite or probable, and 126 more as possible. It is expected that at least half of the latter are actual craters. If the crater density in the areas covered by pictures 7 through 12 is taken to be typical of the whole planet, then there should be somewhere between 60 000 and 100 000 craters, with diameters exceeding 5 kilometers, on the surface of Mars.

In general, the Martian craters resemble those on the Moon in several respects. The size distribution, except perhaps for the smallest and largest, is similar. Some of the craters on Mars have well-defined polygonal, rather than circular, shapes and there are also a number, as on the Moon, with central peaks. On the whole, however, the Martian craters are smoother in outline and have lower walls and shallower interiors than lunar craters.

These differences, taken together with the relatively low frequency of the smaller craters on Mars, suggest that there has been a greater degree of erosion on this planet than on the Moon. The great majority of the smallest observable Martian craters have sharp outlines, whereas the larger ones are usually less distinct. This indicates that many smaller craters have been completely eroded away over the course of time and that most of those now existing were formed relatively recently. The largest craters, on the other hand, which are the most difficult to erode, are much older, on the average.

Possible causes of erosion on Mars are obliteration of smaller craters by larger ones produced subsequently, the transport of particles by the wind, which is known to be capable of attaining hgh velocities, and thermal creep. The latter phenomenon is familiar on Earth; pieces of rock will move down a slope as a result of alternate expansion and contraction arising from temperature changes. The large daily temperature fluctuations on Mars would make thermal creep especially effective.

On Earth, an important source of erosion is flowing water. Although it is highly improbable that there are now streams or bodies of water on Mars, there is no clear proof that they may not have existed in the distant past, during the first 1 or 2 billion years after the formation of the planet. Because of the erosion by other means that has undoubtedly occurred in more recent times, the signs of water flow would have been obliterated.

C. R. Chapman, J. B. Pollack, and C. Sagan wrote in 1968:

> If substantial aqueous-erosion features—such as river valleys—were produced during earlier epochs on Mars, we should not expect any trace of them to be visible on the Mariner IV photographs unless they were of greater extent than typical comparable features on Earth. . . . Thus, any conclusions . . . that the apparent absence of clear signs of aqueous erosion excludes running water during the entire history of Mars . . . must certainly be regarded as fallacious.

The suggestion has been made that some of the Martian craters, and also some lunar craters, have a volcanic, or internal, origin. In 1968, for example, J. A. O'Keefe and P. D. Lowman indicated the possibility that circular areas on Mars may be similar to ring dikes on Earth. The latter are fissures, produced by volcanic action, that became filled with

magma which subsequently solidified. On the whole, however, the general opinion seems to be that the majority of the Martian craters are impact craters. Like most of the craters on the Moon and a few terrestrial craters, they have been formed by the impact of meteoroidal, asteroidal, and cometary fragments during the past 4.6 billion years or so.

Soon after the Martian craters, less than 100 in number, had been identified in the Mariner IV pictures, a rough comparison was made with the average density of craters on the Moon. As a result, the conclusion was drawn by the Mariner IV experimenters, R. B. Leighton and his associates, that—

> the heavily cratered surface of Mars must be very ancient—perhaps 2 to 5×10^9 years old . . . [and] it is difficult to believe that free water in quantities sufficient to form streams or to fill oceans could have existed anywhere on Mars since that time. The presence of such amounts of water . . . would have caused severe erosion over the entire surface.

The foregoing conclusion concerning the age of the craters on Mars was soon called into question by several scientists. They pointed out that, in view of the proximity of the planet to the asteroidal belt, the rate of formation of craters on Mars should have been significantly greater than on the Moon. In comparing the age of the surface of Mars with that of the lunar surface, allowance should be made for this difference. As a result of various calculations it appeared that the average age of the Martian craters was only from 300 to 900 million years. Since the age of the planet is some 4.6 billion years, it followed, in the words of E. Anders and J. R. Arnold, that "the crater density on Mars no longer precludes the possibility that liquid water and a denser atmosphere were present . . . during the first 3.5 billion years of its history."

In the course of recent years, it has become clear that any estimate of the age of the Martian surface based on crater densities involves so many uncertainties and requires so many assumptions that the results have little value. There is a general feeling, however, that many of the larger craters on Mars have an average age of more than 2 billion years, so some are almost as old as the planet itself. The smaller craters, however—those less than 30 kilometers (19 miles) in diameter—which constitute over 80 percent of those with diameters exceeding 5 kilometers, may have ages of only some tens or hundreds of million years. Thus, it appears justified to conclude that the Mariner IV photographs provide no evidence, one way or another, concerning the possible existence of bodies or streams of water in the early history of Mars.

MARTIAN SURFACE TEMPERATURES

Calculated Temperatures

An approximate idea of the maximum and average temperatures of the surface of Mars can be obtained by supposing that the planet behaves as an ideal (blackbody) emitter of radiation. The rate E at which energy is radiated from a blackbody is related to the absolute temperature T in °K, which is 273° plus the temperature in °C, by the theoretical (Stefan-Boltzmann) expression

$$E=8\cdot28\times10^{-11}\,T^4\ \text{cal/cm}^2\text{/min}$$

If it is assumed that the planet has attained temperature equilibrium with its surroundings, as is highly probable, then energy is emitted as radiation at the same rate as it is being absorbed from the Sun. The quantity E can then be taken to be equal to the rate of absorption of solar radiation.

At its average distance from the Sun, the surface of Mars absorbs solar radiation at the rate of 0.66 cal/cm²/min (p. 66). Hence, in the equation given above, E may be set equal

to 0.66. The value of T is then found to be 299° K; that is, 299° − 273° = 26° C (or 79° F). If Mars does not behave as an ideal radiator, as may well be the case, the calculated temperature would be a few degrees higher. This temperature refers to the surface of the planet at its average distance from the Sun. At perihelion, when Mars is closest to the Sun, the calculated temperature would be some 15° K still higher and it would be lower by roughly the same amount at aphelion, when Mars is farthest from the Sun.

The temperatures calculated in this manner represent those at the subsolar point of Mars, i.e., the point where the surface is intersected by a line joining the centers of Mars and the Sun. At this point, at the center of the Martian disk at full phase, not only is the surface closer to the Sun than at any other point but the solar radiation comes in directly from above. At other points, the incoming radiation strikes the surface at an angle, so that a given amount of energy is spread over a wider area. Thus, the blackbody temperature of 299° K, for the average Mars-Sun distance, is the maximum that could be expected, apart from possible allowance for nonideal emission of radiation. Such measurements as have been made, to be described shortly, give maximum temperatures in general agreement with the calculated value.

An average surface temperature can be estimated by noting that solar radiation is absorbed only on the part of the planet that is facing the Sun, over an area of πR^2, but it is lost from the whole surface area $4\pi R^2$, where R is the radius of Mars. Because the temperature is proportional to the fourth root of the radiation absorbed (or emitted), it follows that

$$\frac{T_{\max}}{T_{\text{av}}} = \sqrt[4]{\frac{4\pi R^2}{\pi R^2}} = \sqrt{2}$$

or

$$T_{\text{av}} = T_{\max}/\sqrt{2}$$

where T_{av} is the average temperature over the whole planet and $T_{\max}$ is the value calculated above. If $T_{\max}$ is taken as 299° K, then T_{av} is 211° K; that is, −62° C (−80° F). This temperature also applies when Mars is at its average distance from the Sun. At perihelion it would be about 10° K higher and at aphelion about the same amount lower.

With a maximum daytime temperature of approximately 26° C and an estimated average of roughly −62° C over the whole planet, it is obvious that surface temperatures well below −100° C may be expected on some parts of Mars, especially at night. During the local winter, the polar region does not receive any sunlight for periods up to half a Martian year, so the temperatures must drop to very low values.

Temperature Measurements

Instead of assuming that the rate of emission of energy by Mars is equal to the average rate of absorption from solar radiation, the equation given earlier can be used to calculate the temperature based on an actual determination of the rate at which the planet emits energy. This procedure has the advantage that it gives the temperature at the particular time, and even at a particular part of the Martian surface, for which the energy measurement is made. In calculating the temperature, it is assumed, as a first approximation, that the planet is a blackbody radiator. Then some allowance can be included for nonideal behavior.

The radiation from Mars as received on Earth consists of two parts: first, reflected or scattered sunlight from the planet's disk, and second, thermal radiation from the surface. It is the latter radiation which is dependent upon the surface temperature and is that required for calculating its value. The separation of the thermal radiation from the total

radiation received has been based on the fact that the thermal radiation is essentially all of long wavelength (more than 3μ), and, therefore, is in the infrared region of the spectrum. The scattered sunlight, on the other hand, is nearly all of shorter wavelength.

The experimental technique originally used for determining the thermal radiation was to measure the total energy received in a given time from Mars by means of a heat-sensitive device, such as a vacuum thermopile. Another measurement of the energy of the same radiation after it had passed through a layer (cell) of water or a sheet of glass was then performed. The water or glass absorbed the infrared (thermal) radiation, but allowed the scattered sunlight to pass through. The difference between the two measurements then gave the value for the thermal radiation only. By substituting this quantity for E in the equation on page 131, the corresponding blackbody temperature was calculated.

The earliest radiation measurements for the purpose of determining the temperature of the surface of Mars were made during the 1922, 1924, and 1926 apparitions by W. W. Coblentz and C. O. Lampland at the Lowell Observatory and by E. Pettit and S. B. Nicholson at the Mount Wilson Observatory. As a general rule, the water-cell transmission method was employed to separate the thermal from the total radiation. Although the results indicated certain definite trends in the temperatures, some of which will be described later, there is doubt concerning the absolute accuracy of the values reported.

An improved technique for determining Martian surface temperatures was developed by W. M. Sinton and J. Strong during the 1954 apparition, utilizing the 200-inch telescope at Mount Palomar. They did not employ the water-cell method, because it is susceptible to considerable error, especially because it involves the difference between two small quantities. Instead, they used filters to restrict the radiation energy received to the infrared wavelength range of 7 to 13 μ. Furthermore, the energy was measured by means of gas expansion in a Golay detector, which has about twice the sensitivity of a thermopile.

The results near the Martian equator obtained by Sinton and Strong, during the terrestrial nights of July 20 and 21, 1954, are indicated by the points in figure 6.21. The curve shown is based on theoretical considerations; it gives estimated temperatures during the Martian day as well as the night when no measurements are possible from Earth. As on Earth, the maximum temperature during the day is reached an hour or so after noon. This maximum is close to 300° K (27° C), and hence agrees well with the calculated maximum temperature at the subsolar point.

The estimated temperatures at night, even at the equator, are seen to be quite low, perhaps as low as 190° K (−83° C). The very large daily change in temperature, about 110° C, at the equator, is the result of the low atmospheric density and the poor thermal conductivity of the surface material. Consequently, a considerable amount of heat energy

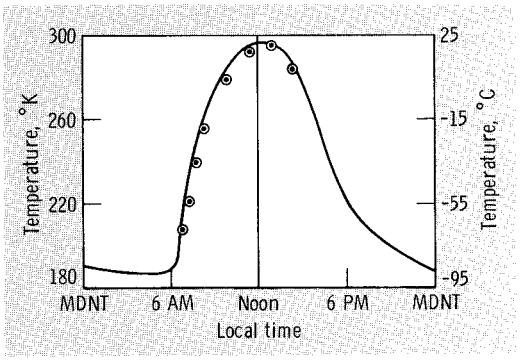


FIGURE 6.21. Daily temperature variation at the Martian equator. (After W. M. Sinton and J. Strong.)

is radiated into space at night, without any appreciable replacement from the subsurface or the atmosphere.

The results of an analysis of the data of Sinton and Strong made by C. Leovy in 1965 imply that the conditions on the Martian surface are such that the daily temperature variations are determined mainly by the size of the surface particles, but not by their nature, and by the low atmospheric pressure. The conclusion drawn is that the particles must be quite small, not more than several microns in diameter.

An indication of the manner in which the daily maximum (noon), minimum (sunrise), and mean surface temperatures vary with latitude is provided by the curves in figure 6.22. These were prepared by the meteorologist Y. Mintz, University of California, Los Angeles, partly from experimental data and partly from calculations. The temperatures are those at the time of the winter solstice (midwinter) in the northern hemisphere and the summer solstice (midsummer) in the southern hemisphere. Thus, the curves show approximately the highest and lowest surface temperatures attained during the Martian year. The temperatures given may not be exact, but the general trends are probably reasonably correct.

It should be remembered that the large daily changes in temperature—roughly 100° C or so at the equator, but less at higher latitudes—are those at the surface. By analogy with terrestrial conditions, it is expected that, during the daytime, temperatures in the atmosphere a foot or two above the surface may be lower than the ground temperature by 20° to 30° C (or more). At night, the difference is probably not as great and may become quite small by sunrise. In any event, it is thought that, as on Earth, the daily temperature variation is considerably less in the atmosphere than it is at the actual surface.

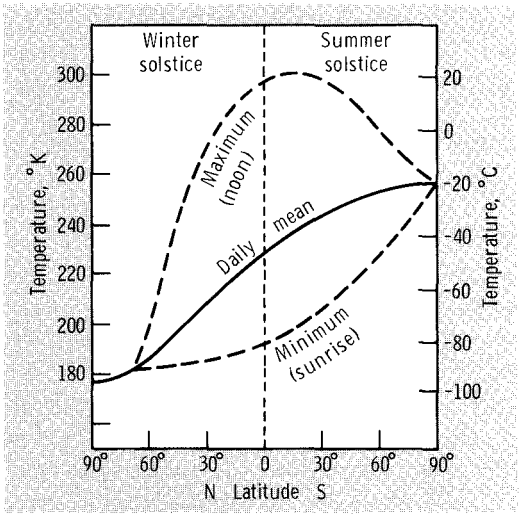


FIGURE 6.22. Temperature variations with latitude at Martian solstices. (After Y. Mintz.)

A modification of the infrared radiometric method for determining the temperature of a planet is to measure the radiation emitted at longer wavelengths, especially in the microwave (centimeter wavelength) region of the spectrum. Although the energy intensities are fairly weak, they can be determined with fair accuracy. Furthermore, at long wavelengths there is a simple relationship between the energy received and the temperature of the emitting body. For a given long wavelength, the absolute (Kelvin) temperature is directly proportional to the radiated energy.

During the apparitions of 1956 and 1958, C. H. Mayer and his associates of the U.S. Naval Observatory studied the emission from Mars at a wavelength of about 3.15 centimeters. The average blackbody temperature, over the whole disk of the planet, was found to be approximately 210° K (−63° C). In subsequent years, other radio astronomers have made measurements at wavelengths up to 21.3 centimeters. Although there are some variations in the temperatures derived, they are

nearly all within the range of 200° to 220° K, with a rough average of around 210° K.

There is a significant difference between the temperature (210° K) estimated from measurements at the longer (microwave) wavelengths and the mean infrared daytime temperature (about 250° K) over the whole disk. The reason generally accepted for this difference is that the infrared temperature is that of the actual surface of Mars, but the radiations of longer wavelength originate from a short distance below the surface. Because of the poor thermal conductivity of the surface material, the subsurface temperatures do not change significantly from day to night. It is probably more than coincidence that the presumed subsurface temperature of about 210° K is very close to the average temperature (211° K) over the whole planet calculated earlier. Such agreement is to be anticipated.

General Conclusions

The following general conclusions may be drawn from measurements and calculations relating to the surface temperature on Mars.

The maximum temperatures in the Martian equatorial regions, about 25° C (77° F), are not much lower than those reached in the temperate zones of Earth. During the summer, temperatures in the daytime are above the freezing point of water, 0° C (32° F), probably at all latitudes up to about 60° N (in the northern summer) and 70° S (in the southern summer).

Maximum surface temperatures on Mars are attained about an hour after midday and they decrease rapidly, at first, and then more slowly after sunset. During the night, there is no part of the planet where the temperature is above 0° C or even above −20° C (−4° F). The lowest surface temperatures are experienced just before sunrise, after which they increase rapidly toward the maximum. The daily surface temperature range may be more than 100° C (180° F) near the equator, but it is less at higher latitudes. The variation is expected to be smaller in the atmosphere just above the surface than it is at the surface.

As a result of the very low nighttime temperatures, a large proportion of the water vapor in the atmosphere will condense out as hoarfrost before sunrise in the winter months. It will vaporize again, however, and possibly melt temporarily into liquid water in certain locations, during the daytime.

The surface of the dark areas of Mars is about 10° C (18° F) warmer than adjacent bright areas in the daytime. This is because of the lower albedo of the former. The dark areas reflect less of the solar radiation and so they absorb more heat from the Sun than do the bright areas. The surface temperatures are consequently higher in the dark areas in the daytime, but the difference is expected to be small at night.

Temperatures just below the surface are around −60° C (−76° F) and so are well below the freezing point of water. In general, the subsurface temperatures are probably lower in the daytime and higher at night than they are at the actual surface. The atmosphere, a foot or so above the ground, may be colder than the surface, especially in the daytime. The manner in which the atmospheric temperature is thought to vary at higher altitudes was described in chapter V.

The surface temperatures during the local summer are higher in the southern than in the northern hemisphere. The winter temperatures are probably lower, as stated in chapter III. The temperatures rise rapidly as summer approaches, and the highest of the year are attained a short time after the local summer solstice. Similarly, the lowest surface temperatures are experienced soon after the winter solstice. The delay between the re-

spective solstices and the maximum and minimum temperature are expected to be less than on Earth because of the low-density atmosphere and the absence of bodies of water.

FOR FURTHER INFORMATION

ADAMS, J. B.: Lunar and Martian Surfaces: Petrologic Significance of Absorption Bands in the Near-Infrared. Science, vol. 159, 1968, p. 1453.

CHAPMAN, C. R.; POLLACK, J. B.; AND SAGAN, C.: An Analysis of the Mariner 4 Photographs of Mars. Smithsonian Astrophys. Obs. Spec. Rept. 268, 1968.

DOLLFUS, A: Polarization Studies of Planets (ch. 9). Visual and Photographic Studies of Planets (ch. 15). Vol. III of The Solar System: Planets and Satellites, G. P. Kuiper and B. M. Middlehurst, eds., Univ. of Chicago Press, 1961.

FOCAS, J. H.: Seasonal Evolution of the Fine Structure of the Dark Areas of Mars. Planetary and Space Sci., vol. 9, 1962, p. 371.

HESS, S. L.: Mars as an Astronautical Objective. Advan. Space Sci. Technol., vol. 3, 1961, p. 151.

LEIGHTON, R. B.; AND MURRAY, B. C.: Behavior of Carbon Dioxide and Other Volatiles on Mars. Science, vol. 153, 1966, p. 136.

LEIGHTON, R. B.; ET AL.: Mariner Mars 1964 Project Report: Television Experiment, Part I: Investigators' Report. Jet Propulsion Lab. Tech. Rept. 32–884, pt. I, 1967.

LEOVY, C.: Mars Ice Caps. Science, vol. 154, 1966, p. 1178.

LOOMIS, A. A.: Some Geologic Problems of Mars. Geol. Soc. Am. Bull., vol. 76, 1965, p. 1083.

MICHAUX, C. M.: Handbook of the Physical Properties of the Planet Mars. NASA SP–3030, 1967.

MOROZ, V. I.: Physics of Planets. Ch. II, NASA TT F–515, 1968.

O'CONNER, J. T.: Mineral Stability at the Martian Surface. J. Geophys. Res., vol. 73, 1968, p. 5301.

O'LEARY, B. T.; AND REA, D. G.: Mars: Influence of Topography on Formation of Temporary Bright Patches. Science, vol. 155, 1967, p. 317.

ÖPIK, E. J.: Atmosphere and Surface Properties of Mars and Venus. Prog. Astronaut. Sci., vol. 1, 1962, p. 261.

ÖPIK, E. J.: The Martian Surface. Science, vol. 153, 1966, p. 255.

POLLACK, J. B.; AND SAGAN, C.: Secular Changes and Dark Area Regeneration on Mars. Icarus, vol. 6, 1964, p. 434.

POLLACK, J. B.; AND SAGAN, C.: Analysis of Martian Photometry and Polarimetry. Smithsonian Astrophys. Obs. Spec. Rept. 258, 1967.

POLLACK, J. B.; GREENBERG, E. H.; AND SAGAN, C.: Statistical Analysis of the Martian Wave of Darkening and Related Phenomena. Planetary and Space Sci., vol. 15, 1967, p. 817.

Proceedings of the Caltech-JPL Lunar and Planetary Conference. Jet Propulsion Lab. Tech. Mem., no. 33–266, 1966, pp. 207–307.

SAGAN, C.; AND KELLOGG, W. W.: The Terrestrial Planets. Ann. Rev. Astron. Astrophys., vol. 1, 1962, p. 235.

SAGAN, C.; AND POLLACK, J. B.: A Windblown Dust Model of Martian Surface Features and Seasonal Changes. Smithsonian Astrophys. Obs. Special Rept. 255, 1967.

SAGAN, C.; AND POLLACK, J. B.: Elevation Differences on Mars. J. Geophys. Res., vol. 73, 1968, p. 1373.

SALISBURY, J. W.: The Light and Dark Areas of Mars. Icarus, vol. 5, 1966, p. 291.

SALISBURY, J. W.; AND HUNT, G. R.: Martian Surface Materials: The Effect of Particle Size on Spectral Behavior. Science, vol. 161, 1968, p. 365.

SINTON, W. M.; AND STRONG, J.: Radiometric Observations of Mars. Astrophys. J., vol. 131, 1960, p. 459.

SINTON, W. M.: On the Composition of the Martian Surface Material. Icarus, vol. 6, 1967, p. 222.

SLIPHER, E. C.: The Photographic Story of Mars. Sky Publishing Corp., 1962.

TOMBAUGH, C. W.: Evidence That the Dark Areas on Mars Are Elevated Mountain Ranges. Nature, vol. 209, 1966, p. 1338; reply by Wells, R. A.: *ibid.*, vol. 209, 1966, p. 1338.

VAUCOULEURS, G. DE: Physics of the Planet Mars. Faber & Faber, 1954.

VII

The Clouds and Haze of Mars

THE MARTIAN CLOUDS

ON THE WHOLE, the atmosphere of Mars is remarkably clear, especially to light of medium and long wavelengths; that is, from green to red and infrared. If this were not so, the telescopic study of the Martian surface would be much more difficult than it already is. Nevertheless, clouds do appear from time to time, and often persist. The visible Martian clouds are classified according to their color: white, blue-white, blue, and yellow.

The color of a cloud, which actually represents the part of the sunlight reflected by the cloud, is best determined with the aid of color filters. A blue filter, for example, absorbs yellow and red light and permits the passage of radiations of short wavelength at the blue end of the spectrum. The blue clouds are those which appear especially bright to the eye or to a photographic film when viewed through a blue filter. White and blue-white clouds are also visible through a blue filter, but the yellow clouds are not. The yellow clouds, on the other hand, can be seen through a yellow filter, and they can also be detected by the eye without a color filter.

In spite of their apparent color variation, it is not certain that there is any fundamental distinction between the white, blue-white, and blue clouds. The same or similar clouds have sometimes been described as white by one observer, but as blue or blue-white by another. The difference in appearance may depend on the size of the particles that constitute the cloud or haze and on its opacity, i.e., by the particle density or number of particles per unit volume.

The distinction between white and blue clouds on Mars is based partly on the polarization measurements made by A. Dollfus. He reported in 1957 that—

> "the shape of the [polarization versus phase angle] curve [for the white clouds] is radically different from that of a cloud of [liquid] droplets, but similar to that for a fog of ice particles. The white clouds . . . must therefore consist of veils of [small] crystals like cirrus clouds [in Earth's atmosphere].

On the other hand, the polarization of the blue clouds, or veils, indicated that "they are not ice crystals but small droplets, presumably

[of] water." It is very difficult, however, to accept the idea that liquid droplets, whether they be of water or of carbon dioxide or of any other substance that might reasonably be present, could form and remain for any length of time in the Martian atmosphere. Liquid water might occur in local warm places on the surface of the planet, as indicated in chapter VI, but it seems very unlikely that the proper conditions would ever exist in the Martian atmosphere.

The view that the white clouds consist of solid particles of water (ice) or of carbon dioxide is widely accepted, simply because it is not conceivable that they can be anything else. For the same reason, it seems probable that the blue and blue-white clouds would be made of similar particles, although of a different size. It is not always possible to draw unequivocal conclusions from polarization measurements and, furthermore, it is well known that the variation of the polarization of particulate matter with phase angle is often strongly dependent on the size of the particles.

Observers of Mars have noted that the blue veils are frequently related to the more dense white-cloud formations, with the blue clouds at higher elevations. For this reason, and also because they cannot always be distinguished from one another and because they probably have basically the same constitution, the so-called white, blue-white, and blue clouds or veils will be considered here as if they are identical. They will be referred to here only as white clouds.

There is no question about the distinctive nature of the yellow Martian clouds. They are quite different in origin, appearance, behavior, occurrence, and constitution from the other clouds. The yellow clouds are clearly apparent with yellow, orange, and red filters, and also without any filter at all. They are not seen with a blue filter because it absorbs the visible light of long wavelength, from yellow through red.

In addition to the visible white and yellow clouds, it is possible that the Martian atmosphere contains a haze which is not visible to the eye or detectable by a photographic film. Although this haze, if it exists, is invisible, it is incorrectly referred to as the "blue haze" or "violet layer." It is certainly not blue or violet in color, because it does not reflect sunlight in the short-wavelength region of the spectrum. The blue haze behaves as if it has the property of absorbing blue light reflected from the surface of Mars. Consequently, if it reflects sunlight at all, which appears doubtful, it must be in the yellow and red wavelength regions.

The White Clouds

White spots on the disk of Mars, that were evidently clouds, were first reported by P. A. Secchi in 1858 and by W. R. Dawes in 1864. The identity of these spots as clouds was recognized by N. Green in 1877. He found that they appeared only on the limb of Mars and remained at the limb as the planet rotated. Because they did not move with the planet, the white spots could not be on the surface. They must have been high in the atmosphere and so the conclusion was drawn that they were clouds.

Further confirmation of the presence of white clouds in the Martian atmosphere was reported in 1890 by J. E. Keeler in the United States and, at the next apparition in 1892, by J. Perrotin in France. These astronomers noted that the white spots extended beyond the terminator of Mars, and so they must be clouds at a considerable height above the surface. From observations made of the extension of the clouds illuminated by the Sun, as seen beyond the terminator, that is, within the shadowed part of the Martian surface,

the height of the white clouds was estimated to be about 5 to 7 kilometers (about 3 to 4.5 miles).

The formation of the white clouds on Mars, like the formation of clouds on Earth, follows a relatively complex pattern. Nevertheless, some conclusions can be drawn on the basis of studies made at all oppositions since the end of the 19th century. White clouds may be divided into two broad categories, depending on whether their formation is determined primarily by the time of day on Mars or by the location on the planet.

During the Martian local spring and summer, in particular, when the polar cap in a given hemisphere is receding, a light cloud or haze formation is often seen at the morning edge of the planet. This edge is at the right (west) of the inverted image as seen in a telescope. It has just emerged into daylight at the end of the night as a result of the rotation of Mars about its north-south axis.

The morning cloud remains stationary, in spite of the rotation of the surface beneath it, and it extends for only a short distance across the planet. This means that the cloud or haze, which presumably formed during the Martian night, where it cannot be seen, is dispersed during the first few hours of the morning. The cloud formation is not confined to any particular locality, but ocurs at each place, generally at low latitudes, not too far from the equator, at about the same local time.

The size and intensity of the morning haze, which has been compared to tropical fog, varies, presumably depending on local conditions, such as the amount of water or carbon dioxide in the atmosphere and the temperatures at different levels. In some regions the haze may appear several days in succession, whereas in others it is seen only occasionally. It seems probable that the haze or cloud forms during the night at an altitude of a few miles where the temperature is low and it disperses, by sublimation of the solid particles, in the early-morning hours as the Martian atmosphere is warmed up by the Sun.

A haze similar to that seen at the sunrise limb is often observed to form at the other limb, just before the Martian sunset. Presumably this results from a process of condensation in the upper atmosphere as it cools in the early evening when the Sun is going down. From observations made on 140 nights during the apparition of 1964–65, C. F. Capen and his associates reported morning haze or cloud on two-thirds and evening haze on about half of the Martian days. In some instances the successive evening and morning hazes were observed at the same latitude, and so it is possible that the cloud formation persisted all through the night and was not dispersed until some time the following morning.

In the second general class of white clouds, where the formation depends on the locality, there are actually two distinct types. One is the polar hood or polar haze that covers the polar cap during most of the local autumn and winter. The hood commences to form in the autumn and, although variable, it persists almost continuously over the polar region until it disperses in a few days just before the spring equinox in each hemisphere. When the haze disappears, it is seen that the polar hood has extended some distance beyond the actual cap which is now visible. The cap is more brilliantly white than the hood that covered it.

The polar hood probably consists of minute crystals, suspended high in the atmosphere, of the same material, either water or carbon dioxide, or both, that constitutes the cap. Toward the end of winter, as the polar region begins to experience more hours of daylight, the increase in temperature of the atmosphere causes the solid particles in the cloud to vaporize. Consequently, the hood disappears. The process is reversed at the beginning of the

319-889 O—69——5

following autumn. The short days and long nights cause the atmosphere to cool once again and the polar hood is formed.

The other type of localized white cloud develops in the late afternoon over certain regions at low and medium latitudes. By sunset, these clouds are large and bright. They commonly have dimensions of roughly 1000 by 1000 kilometers (about 620 by 620 miles), so they cover an area of something like a million square kilometers (380 000 square miles). These large cloud formations are not usually seen on the morning limb the following day, as they evidently disperse during the night. Clouds of this type generally recur over the same region day after day, so they have been called recurrent clouds. During the 1964–65 study of Mars, recurrent cloud activity was seen on over one-fifth of the observing days. These were mostly in the late Martian spring season when one polar cap was diminishing and the other increasing in size at the maximum rate.

Since 1907, it has been known that certain areas on the surface of Mars seem to be conducive to the development of the recurrent afternoon white clouds. Furthermore, once the cloud has formed completely, it remains almost stationary, apart from some drift caused by the wind, over the area where it formed. The afternoon clouds thus follow the rotation of the planet. This clearly distinguishes these clouds from the morning and evening clouds, or haze.

Among the regions over which recurrent clouds often form are the following: Tharsis (between Candor and Phoenicis Lacus), Elysium, Hellas, and localities in the Arcadia-Amazonis area (for example, Nix Olympica). All these are bright regions, and the recurrent generation of essentially stationary clouds over them has led to the belief, based on the terrestrial analogy, that they are elevated areas. Reasons for thinking that this may perhaps not be so have been given in chapter VI.

A curious cloud formation of the recurrent (afternoon) type, called the W-cloud group, was described by E. Pettit and R. S. Richardson at the Mount Wilson Observatory in 1954. It appears, however, that the formation was observed in 1907 by E. C. Slipher, who claimed to have seen it in the Tharsis region, near Phoenicis Lacus, on several, but not all, Martian apparitions since that year. In 1954, the group of clouds, shaped somewhat like the letter W, in the usual telescopic view of Mars with the south pole at the top, was seen to form every afternoon for a period of more than 3 weeks. But during the next opposition in 1956, the W-formation could not be observed. It will be seen shortly that the period of the 1956 apparition was also exceptional in another respect. The W-cloud reappeared, though, in 1958.

An unusual aspect of the W-cloud group is that, although it occurs over a generally bright area, the individual components of the W seem to coincide with well-known canals and oases, which are dark regions. In writing about his observations, R. S. Richardson said: "The strokes of the W were formed by narrow stripes about 1000 miles in length. There were knobs at the top or western end of the W and at the three intersections of the stripes." The knobs appear to fall on or close to known oases, whereas the stripes of the W coincide with the canals Draco, Fortunae, and Ulysses.

What may be a kind of recurrent localized mist formation has been observed, mainly during the period between late spring and early summer in each hemisphere. Bright regions, such as Hellas, Isidis Regio, and Elysium, which are normally light ocher in color and are surrounded by dark areas, are seen to be covered with a white layer during the morning. It has not been established, however, if this layer is a low-level, ice-fog type of phe-

nomenon or if it is surface frost, either of water or carbon dioxide.

The Yellow Clouds

The formation of clouds having "the color of Mars" was suggested by H. Flaugergues in 1809, and it was reported by C. E. Burton in Ireland in 1879, by W. H. Pickering in 1894, and by A. E. Douglass in 1899. These yellowish clouds range in opacity from light veils, which only partially obscure or soften the outlines of surface features, to opaque clouds, sometimes called duststorms, which cover large areas of the Martian surface. The yellow clouds seem to originate in the bright areas and they are seen both on the disk of the planet and as projections beyond the edge. They are relatively uncommon and none were reported by C. F. Capen in the 1964–65 apparition.

Because of their color and origin, it is generally accepted that the yellow clouds are very fine particles of the material covering the bright areas carried aloft by the wind. The difference in particle size would account for the color of the clouds being somewhat more yellowish than the ground. The polarization of the yellow clouds is not quite the same as that of the surface, but this is not unexpected. Both the particle size and the suspension in the atmosphere can affect the polarization.

The less-dense yellow clouds generally last for a day or two, and they travel, presumably driven by winds, at average speeds which are usually from 15 to 50 km/hr (10 to 30 mph). On occasions much higher speeds have been reported, especially in the early stages of development of the larger cloud systems (see below). It appears that the yellow clouds often have a tendency to move around the edges of dark areas and to avoid their centers.

There is probably a continuous gradation between the yellow clouds described above, covering a limited area, and the extensive and dense duststorms. The latter are so distinctive, however, that they merit separate consideration. The intense duststorms, which blot out large areas of Mars for many days and often weeks at a time, are quite rare. They seem to originate in a particular bright area and to extend both in latitude and longitude.

Toward the end of August 1956, for example, a yellow cloud was seen to form near Argyre and to spread north and south. It drifted westward until it formed a band almost completely surrounding the planet at middle latitudes in the southern hemisphere. The south polar cap was also covered and remained hidden until the middle of September. This Martian duststorm, which occurred during the apparition of 1956, was one of the most severe on record and greatly interfered with observations of the planet.

In 1909, E. M. Antoniadi pointed out that, although duststorms occur at other times, the more intense storms seem to develop during the perihelion oppositions (ch. III). This idea obtained support from observations made in later years, when severe storms occurred at the subsequent perihelion oppositions in 1924, 1941, and 1956. Because Mars is closest to the Sun at perihelion, there is a maximum heating of the surface. Strong convection currents and turbulence in the atmosphere might then be expected to generate violent winds which will raise fine particles of dust from the surface and carry them great distances.

Windspeeds are determined by estimating the rate of motion of the dust cloud across the surface of the planet. At the beginning of a duststorm the windspeed is often quite high. For example, in 1967 G. de Vaucouleurs stated that "velocities of up to 100 km/hr [over 60 mph] have been recorded for only a few hours." The speed of the wind then

decreases and after the first day it has dropped to around 60 km/hr (37 mph). Subsequently, the windspeed falls off more slowly.

There is some evidence that duststorms occur on Mars in a particular hemisphere toward the end of spring when the polar cap in that hemisphere is receding at its fastest. This has led some scientists to suggest that there might be a connection between the development of duststorms and water vapor. The amount of the latter in the atmosphere certainly increases in the late spring, but it is possible that the same external factors are responsible for both the rapid disappearance of the polar cap and the duststorms. There might then not necessarily be any direct relationship between the two phenomena.

There appears to be some difference of opinion concerning the heights in the atmosphere to which duststorms can extend. Thus, in referring to the great storm of 1956, E. C. Slipher said that "blue and yellow photographs [i.e., photographs taken with blue and yellow filters] show that the dust clouds were higher than the blue-white polar winter cloud cover." On the other hand, A. Dollfus reported that "in the polar and temperate regions, well-defined white clouds were observed above the [dust] veil." When illuminated by the setting Sun and seen as projections beyond the terminator of Mars, intense duststorms appear to reach altitudes of up to about 30 kilometers (19 miles). There is reason to believe, however, that not all yellow clouds attain such heights. Many probably do not get much above an altitude of about 6 kilometers (3.7 miles).

After a duststorm, those dark areas which have been traversed by the yellow cloud or clouds have a somewhat lighter color, presumably as a result of deposition of light-colored particles. But within 2 or 3 weeks the original appearance of the dark areas is restored. As mentioned in chapter VI, this behavior has been regarded as favoring the theory that the dark areas support some kind of vegetation, although it is not necessarily so. Because they are normally very bright, the polar caps are seen with reduced brightness after the passage of a duststorm.

A curious event was observed after the duststorm of 1956. At the time this storm occurred, it was late spring in the southern hemisphere and the polar cap had almost

Orange Yellow Blue

FIGURE 7.1 Photographs of Mars taken with orange, yellow, and blue filters. (Lowell Observatory photographs.)

entirely disappeared. But when the yellow duststorm cloud lifted, a "new . . . brilliant white [polar cap] uncontaminated by the atmospheric dust" was observed, according to R. S. Richardson, "just before the advent of summer." Presumably, the yellow dust had absorbed a sufficient amount of the Sun's heat radiation to cause a significant lowering of the surface temperature. As a consequence, water or carbon dioxide would then be deposited from the atmosphere in solid form.

The suggestion was made in 1960 by C. C. Kiess and his associates that the yellow Martian clouds do not consist of dust particles but of nitrogen dioxide (NO_2) gas. This is an aspect of the unsubstantiated theory mentioned in preceding chapters that the atmosphere of Mars contains significant amounts of this gas and of its dimeric form nitrogen tetroxide (N_2O_4). An increase in atmospheric temperature would favor the formation of the brown dioxide from the much lighter colored tetroxide. In the opinion of the proponents of the theory, it would then appear as if a yellow storm had developed. It has already been mentioned, however, that if there is any nitrogen dioxide (or tetroxide) in the atmosphere of Mars, the quantity is extremely small. The amount, if any, present would be quite inadequate to account for the observations.

THE BLUE HAZE

General Characteristics of the Blue Haze

When taking photographs of Mars through filters of various colors, C. Lampland found in 1909 that there was a tendency for the pictures to be indistinct when blue filters were used. It has since been established that telescopic images with yellow, orange, red, and infrared filters show the familiar surface details, but the details disappear when blue, violet, and ultraviolet filters are used. The polar caps, however, generally appear brighter in blue or violet light (fig. 7.1). There seems to be a fairly definite wavelength in the blue region of the spectrum, about 4500 Å, below which the surface details on Mars, that is, the separations between bright and dark areas, are no longer apparent. Except for the polar caps, the surface of the planet has a uniformly blank appearance.

The phenomenon described above has been regarded as being due to a blue haze or violet layer. Actually, there is no visible haze or layer, and if it could be seen it would not appear to be blue or violet in color. Because the surface of Mars shows no detail in light of wavelength less than 4500 Å, the implication seems to be that if the atmosphere is responsible for the effect, then it must contain something, possibly a haze, that prevents passage of blue, violet, and ultraviolet radiations. If this is the case, then the haze could not possibly be blue, but it might have a reddish or orange color. Even this is not certain, as will be seen in due course. Nevertheless, the expression "blue haze," although definitely a misnomer, is still in general use.

Until 1937, astronomers had assumed that the blue haze was a permanent feature of the planet Mars. In that year, however, E. C. Slipher found that there were "rare intervals [when the atmosphere clears] . . . sufficiently to permit the short wavelengths of light [from sunlight] to penetrate to the surface below and come out again." During these rare occasions, now called blue clearings, photographs of Mars taken with blue or violet filters show almost as much detail as is apparent with yellow and red filters.

Sometimes the clearing is partial, the word "partial" being used in two different senses. The clearing may be such that details show up only on certain parts of Mars when photographed with a blue filter, or the details may

not all be equally distinct. In 1962, Slipher reported that "photographs [with a blue filter] taken on the same date at different stations around the world showed wide differences in the opacity of the violet layer, sometimes varying from virtual opacity to nearly complete transparency. . . . Actually, at times, one face of the planet was obscured while the other face was nearly transparent [to blue light]." At other times, however, the clearing is planetwide; all areas on the surface show normal contrast when viewed through a blue or violet filter. When clearing occurs it generally lasts for a few days. Then the apparent opacity to light of short wavelength returns.

At one time, Slipher, who studied the phenomenon over many Martian apparitions, was of the opinion that the blue clearings always occurred at or close to the date of opposition. From an examination, made by two independent observers, of several thousand photographs taken in blue light over a period of a number of years, Slipher concluded, at first, that "the clearings were so striking in character and were so closely associated with the oppositions that such occurrences were accepted to be the rule." No blue clearings were observed, however, at the actual opposition dates in 1954, 1956, and 1958, although there were examples of such clearings within 10 to 14 days of these dates. "All together," wrote Slipher in 1962, "observational evidence tends to support the conclusion that there is [a] definite tendency for pronounced blue clearings to occur more often near opposition. On the other hand, recent studies have disclosed examples of blue clearings as far as 60 or 70 days from [the] opposition date."

In the course of their extended observations of the 1964–65 Martian apparition, C. F. Capen and his associates detected no strong blue clearings, but there were several which were described as "medium." One of these commenced on September 16, 1964, which was 174 days before the opposition date of March 9, 1965. It continued, with some variation in degree, for about 10 days. Another period of medium blue clearing started on December 30, 1964, 69 days before opposition. A similar clearing occurred around the opposition date, and a somewhat weaker one was observed in the middle of July 1965, more than 120 days after opposition.

A survey of the records over a series of Martian oppositions indicates that blue clearings have frequently been apparent over periods of more than 100 days both before and after the opposition date. Nevertheless, the frequency and intensity of the recorded clearings is undoubtedly greater around the time of opposition than at other times. Possible reasons for this situation are, first, that most photographs are taken near the opposition date when Mars is closest to Earth; and second, as Slipher points out, at opposition "the observer looks in a more nearly vertical direction . . . and is provided with the best opportunity to see the surface [of Mars]." Another possibility will be mentioned later.

Interpretation of the Blue Haze: Atmospheric Particles

Several different explanations have been proposed for the blue-haze phenomenon, but there is none so convincing that the others can be rejected. At the present time, therefore, it appears that the choice of a theory is largely a matter of individual preference.

The earliest view concerning the apparent opacity at short wavelengths was that it was the result of a lack of contrast between the light and dark areas. Figure 6.7 shows that the reflectivities of the bright and dark areas become closer and closer together with decreasing wavelength; that is, toward the blue end of the spectrum. For wavelengths of 4500 Å and less, the difference in the re-

flectivities is quite small. Hence in blue or violet light the contrast between the bright and dark areas virtually ceases to exist. Consequently, the boundaries would be very difficult to detect in light of wavelengths below about 4500 Å.

This simple interpretation of the blue haze was abandoned in 1939 when the blue clearing effect was discovered. There seemed no way to account, on the basis of surface reflectivity, for the development of contrast between bright and dark areas and its persistence for a few days. The surface contrast theory has been revived in recent years, however, and it will be discussed more fully at the end of the section on the Martian blue haze.

As an alternative to the explanation of the blue-haze phenomenon as a surface contrast effect, there is the possibility that it is caused by something in the Martian atmosphere. One view, which has been widely held (E. Schatzman, 1952; G. P. Kuiper, 1952; S. L. Hess, 1958), is that very small crystals of ice, about 0.3 to 0.4 micron in diameter, are responsible for the apparent opacity to blue and violet light. Such particles are much smaller than those thought to be present in the visible white clouds.

Very small crystals, formed in a layer high in the atmosphere, can cause forward scattering of the shorter wavelengths in sunlight at the blue end of the spectrum. The longer wavelengths, in the yellow, orange, and red, on the other hand, are scattered in all directions. As a result, details of the Martian surface can be distinguished in yellow and red light, but not in the blue region of the spectrum. The blue clearing could be accounted for by a rise in temperature of a few degrees which would cause the small crystals to vaporize (sublime). By this means, a blue clearing could develop in a very short time.

Because carbon dioxide gas is probably the major constituent of the Martian atmosphere, the blue haze might conceivably consist of very small crystals of solid carbon dioxide. In 1958, S. L. Hess, who originally supported this view, stated that such a layer would be too opaque. Furthermore, it seemed that the temperature conditions in the atmosphere of Mars would lead to the formation of particles of solidified water rather than of carbon dioxide. These conclusions, based on data available at the time, may require modification, however, because of some more recent results on the composition of the Martian atmosphere.

There is a serious objection to the forward-scattering theory of the blue haze, regardless of whether the haze is due to small crystals of ice or solid carbon dioxide. If the particles cause forward scattering of blue light, this component of sunlight will reach the surface of the planet and be reflected. Upon reaching the blue-haze layer, the blue light should again undergo forward scattering, out of the Martian atmosphere toward Earth. Thus, a purely forward-scattering layer above a reflecting surface would be essentially transparent to blue light. If this were so, the surface details of Mars would be just as clear at the blue end of the spectrum as they are at the red end.

In order to overcome this difficulty, it might be supposed that the scattering of blue light from the layer of small particles is predominantly in the backward direction. In this event, the reflectivity (and albedo) of Mars in light of short wavelength, such as blue and violet, would be high, whereas it is actually low. Incidentally, the observed reflectivity in the ultraviolet region is just about what would be expected from the atmospheric pressure alone, and there is no significant contribution that might result from backscattering by particles responsible for the blue haze.

A way out of the dilemma was proposed by G. P. Kuiper in 1961. Although the particles

are assumed to scatter blue light predominantly in the forward direction, he postulated that they also exhibit a small amount of backscattering. Thus, when sunlight falls on Mars, most of the radiation of short wavelength is scattered forward toward the planet, but a small proportion is scattered back to space. The light which reaches the surface of Mars is reflected back toward the particle layer, but because of the poor reflectivity of the surface material for blue light, only a small proportion of reflected blue light will reach the blue layer and be scattered forward to space and to Earth.

The surface reflectivity of Mars for blue light is about 0.05. Hence, the proportion of the blue light that undergoes forward scattering out of the Martian atmosphere to space, after reflection from the surface, is less than about 5 percent of that in the original sunlight. If the proportion of blue and violet radiation in the sunlight that is scattered back directly from the haze layer is more than 5 percent of the total, then it will, according to Kuiper, swamp the blue light coming from the surface. In these circumstances the surface features would be obscured when observed with a blue filter.

An objection based on the nature of the particles (ice or solid carbon dioxide) rather than on their scattering properties had been raised by the American scientist H. C. Urey in 1958. It might be accepted that a purely local blue clearing or restoration of the blue haze could arise from a few degrees' change in temperature high in the atmosphere. It appears improbable that such changes would occur simultaneously over the whole of Mars, as would be necessary to account for clearings (or the reverse) of a planetwide nature. No convincing response has been made to this argument.

Because it appeared that scattering of blue light could not explain the properties of the blue haze in an entirely satisfactory manner, consideration has been given to the possibility that atmospheric absorption of light of short wavelength is responsible. If the postulated haze consists of small, dark-colored particles which absorbed blue and violet light, the only radiations that could be transmitted to space after reflection from the Martian surface would be those of a longer wavelength, yellow and red. Images of the planet in blue light, seen through a blue filter, would then show no surface detail.

In 1953, B. Rosen suggested that the blue haze might be caused by small particles of carbon black, and this view was developed by E. J. Öpik in 1960. Such particles could form in the Martian atmosphere by various reactions following the decomposition of carbon dioxide (or possibly of hydrocarbons) by the ultraviolet radiation in sunlight.

Öpik stated:

> As a working hypothesis, the absorbing substance of the Martian atmosphere may be assumed to consist of two principal components: a blue-absorbing smoke, possibly carbon particles whose concentration is more or less constant; and surface dust, absorbing more evenly over the entire spectral range, whose concentration is variable. The fluctuations in the transmission coefficient [or opacity] may then be caused by the weather-dependent dust content.

Öpik showed that a change of only 25 percent in the amount of dust in the atmosphere could account for the blue clearing. "The blue clearing," he said, "would correspond to . . . [calm] periods when the dust has settled down." At other times, absorption of light by the extra dust in the atmosphere would be responsible for the blue-haze effect.

The following quotation from E. C. Slipher (1962), however, seems to provide a contrary argument to Öpik's views:

> The great yellow cloud which began to appear on August 20 [1956] spread out over most of the planet . . . [and the] conditions continued for about three weeks. During

> this time blue photographs [i.e., taken with blue filters] at Flagstaff [Ariz.], Australia, and elsewhere recorded the maximum blue clearing . . . [whereas] before the storm began, the violet layer [i.e., the blue haze] was practically opaque the whole time.

This observation also appears to dispose of the view, put forward by the English meteorologist R. M. Goody in 1957, that absorption of light by surface dust suspended as very fine particles in the atmosphere was the cause of the blue haze.

One reason Öpik introduced dust into his carbon-particle model of the blue haze was to account for the blue clearings. If the blue haze were caused by small carbon (or other blue-absorbing) particles only, the blue clearing would have to be attributed to the settling of these particles under the influence of gravity. In order to have the required optical properties for the normal blue haze, the carbon particles would be too small to settle fast enough to account for the observed rate of the blue clearing. By postulating that somewhat larger dust particles made a contribution to the light absorption of the blue haze, this difficulty could apparently be overcome.

An alternative solution to the problem of rapid clearing of a blue haze of carbon particles was proposed by E. C. Slipher in 1962. He suggested that the particles might serve as crystallization centers, as small particles frequently do, for atmospheric water vapor or possibly carbon dioxide. If a suitable drop in temperature occurred, the particles would increase in size as a result of the deposition of ice or solid carbon dioxide, and then they would settle out quite rapidly.

The foregoing mechanism would account for the rapidity of the blue clearing, but it is not obvious how the carbon particles would be re-formed in the atmosphere to restore its opacity to blue light. From reported observations, it would appear that recovery of the blue haze may occur even more rapidly than the clearing. Furthermore, Urey's argument, mentioned earlier, concerning the need for a simultaneous change in temperature to cause a planetwide clearing is also applicable in this case.

In another version of the crystallization center theory, Z. Kviz in Czechoslovakia put forward the idea in 1961 that the light-absorbing particles, on which ice or solid carbon dioxide could deposit, consisted of meteoritic dust. This was a development of a suggestion, originally made by F. Link, also in Czechoslovakia, in 1950, that such dust particles, coming from interplanetary space within the solar system, were responsible for the blue haze.

Meteor showers are observed frequently from Earth and some occur at regular annual or biannual intervals. These regular showers are believed to be caused by entry into the atmosphere of small particles which are the debris left behind by existing comets or by the remains of comets that have ceased to exist. The particles form continuous streams orbiting around the Sun. Once or twice every year, Earth passes through each of these streams. At that time a meteor shower is seen. It is made up of many bright streaks (meteor trails) produced by the meteoritic particles as they heat up during their traversal of Earth's atmosphere.

In its orbital motion around the Sun, the planet Mars encounters several streams of meteoritic particles, some of them being the same as those through which Earth passes. In any event, if the Martian blue haze is connected in any way with these particles, there should be some correlation between the blue haze or blue clearing and the location of Mars in its orbit.

For some time, no such correlation was apparent, but in 1965, A. Palm and B. Basu, at the University of California, Berkeley, reported that they found some connection be-

tween blue clearings and the times of minimum meteor-shower activity. These scientists suggested that normally there is a kind of balance between the meteoritic dust particles that enter the Martian atmosphere and those that reach the surface. In these circumstances there are sufficient blue-absorbing particles in the atmosphere to produce the blue haze. When the planet moves out of the meteoritic dust streams, the decrease in the number of particles would cause a blue clearing. Conversely, return to the dust stream would rapidly restore the blue haze.

A comparison has been drawn between the Martian blue haze and the terrestrial phenomenon of noctilucent clouds. These tenuous clouds can sometimes be seen, generally at high latitudes in the summer, glowing faintly in the sky after sunset. Samples of the particles present in the noctilucent clouds, which are at an altitude of some 80 kilometers (about 50 miles), have been collected by means of rockets. Analysis has shown that very small particles 0.05 to 0.5 micron in diameter, are present, and that they contain appreciable amounts of nickel, an important constituent of meteoritic matter. It would seem, therefore, that the noctilucent clouds on Earth are produced by meteoritic dust. There are reasons for believing that in the clouds the dust particles are coated with a layer of ice. Similar clouds might well also exist high in the atmosphere of Mars.

During the Martian apparition of 1956, A. G. Wilson in the United States made a study of the spectrum of the light from the planet in the blue, violet, and ultraviolet region at wavelengths below 4500 Å. Observations were made at times when the blue-haze condition existed and also during blue clearings. In 1958 he reported that the intensities at certain wavelengths in the blue-haze spectrum were greater than they were when the haze cleared. Moreover, these characteristic wavelengths were said to be the same as those found in the spectrum of terrestrial noctilucent clouds, except that the relative intensities are reversed, so that the maxima in the spectrum of the noctilucent clouds corresponds to the minima in the blue-haze spectrum. This difference may perhaps be accounted for by the forward scattering of the sunlight reaching Earth from the noctilucent clouds, whereas the light from Mars might be the result of backscattering or absorption.

If the Martian blue-haze effect is caused by small meteoritic particles, then a similar phenomenon should occur in Earth's atmosphere. Earth should exhibit some opacity in blue light when observed from space at a considerable distance above the atmosphere, but whether this is or is not the case has not yet been determined.

Atmospheric Gases and the Blue Haze

Apart from the idea of surface contrast, the foregoing theories of the cause of the blue haze are all based on the view that solid particles of one kind or another are responsible. In 1957, however, H. C. Urey and A. W. Brewer suggested that the atmosphere of Mars contains certain positive ions, such as CO_2^+, CO^+, and N_2^+, which are known to absorb radiation in the blue and violet regions of the spectrum. The presence of such ions in significant quantities could account for the apparent opacity of the Martian atmosphere to blue light. The Mariner IV studies, described in chapter V, show that the planet has an ionosphere, and this probably contains some CO_2^+ and CO^+ ions, and possibly also N_2^+ ions. But it is not known if the concentrations are sufficient to account for the blue haze.

Blue clearings could be explained by the recombination of the positive ions with the electrons which are present simultaneously.

Subsequent ionization of the neutral molecules would result in a restoration of the blue-haze conditions. Both the recombination and ionization processes can occur rapidly and so there would be no difficulty in explaining the relatively short times in which the blue haze clears or is re-formed. The problem of planetwide clearing, however, to which reference was made earlier, still remains.

In their original presentation, Urey and Brewer proposed that the ions causing the blue-haze effect were produced by the action of high-energy protons (positively charged hydrogen ions) from the Sun. The purpose was to account for the supposed predominance of the blue clearing when Mars is in opposition. At such times, the protons from the Sun must pass close to Earth on their way to Mars. Deflection by the terrestrial magnetic field would then prevent many of the protons from reaching Mars. The normal balance between ionization and recombination would then be disturbed. Recombination would predominate and blue clearing would then result.

A quantitative examination of the situation was made by C. Sagan in 1962. He showed that the rate at which solar protons would have to reach the Martian surface to produce the necessary ionization for the blue haze is extremely high, and that the terrestrial magnetic field would have to extend to unrealistically large distances in order to deflect the protons sufficiently to cause the blue clearing. It should be pointed out, however, that if the supposed correlation between the Martian oppositions and the blue clearing is not exact and can be disregarded, as seems to be possible, then additional sources of ionization can be invoked.

In the terrestrial atmosphere, for example, the ions in the ionosphere, largely NO^+, O_2^+, and O^+, are produced by the ultraviolet and X-radiation from the Sun. Although the average distance of Mars from the Sun is somewhat greater than from Earth, these radiations are still capable of producing considerable ionization in the Martian atmosphere. Any falling off in the solar activity of protons, X-rays, or ultraviolet radiation would be accompanied by a decrease in the ionization and a blue clearing would result. In this event, there should be some parallelism between the activity of the Sun and the blue-haze phenomena.

Another theory of the blue haze, which does not require the presence of solid particles in the atmosphere, is that it is caused by the brownish gas nitrogen dioxide. The brown color indicates that this gas absorbs light very strongly in the blue and violet region of the spectrum. Its existence in the atmosphere in adequate concentration would readily account for the opacity to blue light. Blue clearings would be caused by a decrease in temperature, resulting in the combination of pairs of nitrogen dioxide molecules to form the almost colorless nitrogen tetroxide. As seen in preceding chapters, however, there is no convincing evidence that the atmosphere of Mars contains any of these oxides of nitrogen.

One of the puzzling features of the blue haze has been that, although photographs of Mars in blue and violet light generally show no contrast between bright and dark areas, the polar cap stands out clearly (fig. 7.1). But scattering or absorption of light of short wavelength by constituents of the Martian atmosphere would be expected to decrease the brightness of the caps. It was thought at one time that the material of the polar caps might be exhibiting some kind of luminescence. Spectra of the caps in the short-wavelength region indicate, however, that, like the remainder of the surface, the caps merely reflect sunlight. Explanations of the blue haze based on the absorption or scattering of light (or both) thus encounter a serious difficulty.

The Surface-Contrast Theory

The surface-contrast theory is attracting interest again, following the suggestion made by D. C. Evans in 1965 and the extensive studies reported by J. B. Pollack and C. Sagan in 1967. The latter contend that the blue-haze phenomenon is largely the result of the small difference in the reflectivities of the dark and bright areas of Mars in light of short wavelength, combined with seeing problems which are known to be most severe in blue light. For example, the Wright effect, described in chapter IV, according to which Mars appears to be larger in blue than in red light, might perhaps be accounted for, at least partly, by the poor seeing in the blue region. Furthermore, J. H. Focas has found that when details on the Martian surface can be detected in blue light, that is, when a blue clearing occurs, the dark areas of the surface are distorted in shape and are larger in extent than in yellow light.

If the blue haze is not caused by either absorption or scattering of light of short wavelength by something in the Martian atmosphere, then the brilliance of the polar caps in blue light can be readily understood. The reflectivity of the caps is essentially independent of the wavelength of the light, but the reflectivities of the bright and dark areas fall off markedly with decreasing wavelength. Consequently, the difference in reflectivity between the polar caps and other parts of the surface is greater at short wavelengths, that is, in blue light, than at longer wavelengths, in yellow or red light. As a result, the polar caps will appear to be much brighter, by contrast with the rest of the surface of the planet, in blue than in red light.

According to the view outlined above, the Martian blue haze is largely caused by conditions in Earth's atmosphere, rather than in the atmosphere of Mars. The blue clearings would then correspond to occasions when the seeing in blue light has improved sufficiently to permit the detection of the low contrast between bright and dark areas in this light. Seeing conditions in the terrestrial atmosphere vary from place to place, and they can, and do, change in a short time. In this way it is possible to account for the variability of the blue clearing and the speed with which it can develop or disappear.

"The 1956 observations," wrote E. C. Slipher in 1962, "show plainly that the state of the blue clearing differed widely at Lowell Observatory [Arizona], Lamont-Hussey [South Africa] and Mount Stromlo [Australia], where observations were made on the same night." Such differences are not unexpected if the blue clearing arises from conditions in Earth's atmosphere.

An important contributory factor may have been that the areas of Mars observed were not the same at the three places mentioned above. Because of the poor contrast between bright and dark regions in blue light, it is difficult to discern surface details in such light even under the best seeing conditions on Earth. But when a large dark area, such as Syrtis Major, extending in both east-west and north-south directions, is near the center of the Martian disk, the circumstances are most favorable for seeing details on the surface of the planet (fig. 7.2). Pollack and Sagan state that blue clearings are indeed commonly observed at such times. In the same connection, J. H. Focas (1966) has noted that blue clearings occur most frequently during the seasonal darkening of the dark areas. The contrast between bright and dark areas is then at its greatest. Consequently, the observations of blue clearings will depend not only on the local atmosphere on Earth, but also on the part of Mars that is toward the observer.

Although the conditions for detecting blue clearings could occur at almost any time, they

are most likely to be realized at or near opposition. Mars is then closest to Earth and a given dark area has its maximum apparent dimensions. In these circumstances, some surface contrast, indicating a partial blue clearing, might be observed even when the seeing is only moderately good. Another favorable factor may be that Mars is highest in the sky around the time of an opposition. The light from the planet then passes through a smaller thickness of Earth's atmosphere than at other times.

The Blue Haze and Ultraviolet Radiation

To conclude this section it is still appropriate to quote from a review by C. Sagan and W. W. Kellogg published in 1961. They state that "a definitive explanation of the Martian haze remains an illusive and provocative problem." It is a problem, however, that is of more than mere theoretical or academic interest, because it may have a bearing on the question of whether or not life can exist on Mars. If the blue-haze effect is caused by absorption of light of short wavelength, then the intensity of ultraviolet radiation from the Sun that reaches the surface of the planet would be greatly reduced.

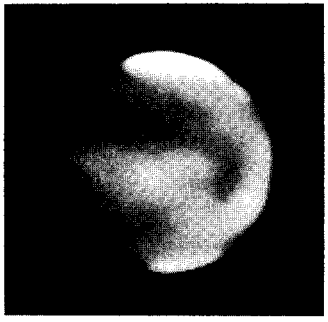


FIGURE 7.2. Mare Tyrrhenum and Syrtis Major regions observed during a blue clearing. (Lowell Observatory photograph.)

Ultraviolet radiation is known to be generally harmful to terrestrial life. The probability of finding life on Mars might thus be decreased if the blue haze were caused by forward scattering or by a surface-contrast effect. On the other hand, it is not impossible that there may exist life forms that are relatively insensitive to ultraviolet light or that have developed some means for protecting themselves from it (ch. X).

In any event, even if the radiations of short wavelength have proved inimical to the development of life, prebiological compounds, that is, chemical compounds from which life may have evolved on Earth, might be found on the Martian surface. According to current ideas, such compounds were probably produced by the action of ultraviolet radiation on the gases in the primitive planetary atmosphere.

FOR FURTHER INFORMATION

DOLLFUS, A.: Visual and Photographic Studies of Planets (ch. 15). Vol. III of the Solar System, Planets and Satellites. G. P. Kuiper and B. M. Middlehurst, eds., Univ. of Chicago Press, 1961.

EVANS, D.: Ultraviolet Reflectivity of Mars. Science, vol. 149, 1965, p. 969.

HESS, S. L.: Mars as an Astronautical Objective. Advan. Space Sci. Technol., vol. 3, 1961, p. 151.

KELLOGG, W. W.; AND SAGAN, C.: The Atmospheres of Mars and Venus. Natl. Acad. Sci.-Natl. Res. Council Publ. 944, 1961.

MOROZ, V. I.: Physics of Planets. Ch. II, NASA TT F-515, 1968.

ÖPIK, E. J.: Atmosphere and Surface Properties of Mars and Venus. Prog. Astronaut. Sci., vol. 1, 1962, p. 261.

POLLACK, J. B.; AND SAGAN, C.: An Analysis of Martian Photometry and Polarimetry. Smithsonian Astrophys. Obs. Spec. Rept. 258, 1967.

SAGAN, C.; AND KELLOGG, W. W.: The Terrestrial Planets. Ann. Rev. Astron. Astrophys., vol. 1, 1962, p. 235.

SLIPHER, E. C.: The Photographic Story of Mars. Sky Publishing Corp., 1962.

VIII

The Origin of Life

HISTORICAL REVIEW

The Theory of Spontaneous Generation

IN CONSIDERING THE POSSIBILITY of the existence of living organisms on Mars, the first step is to examine what is known or, rather, what has been speculated, concerning the origin of life on Earth. Until about 100 years ago, it was commonly accepted that some forms of life could arise spontaneously. This idea is to be found in the records of ancient Babylon and Egypt, as far back as 2000 years before the Christian era, and also later in the writings of the Greek philosophers, including Aristotle.

Living creatures, such as insects and worms, were frequently seen to develop in damp soil, in rotting meat, and in many forms of refuse. The evidence for the spontaneous generation of life consequently seemed to be convincing. In the words of the Russian biologist A. I. Oparin, one of the pioneers of modern attempts to trace the origin of life on Earth: "Among the ancient peoples the belief in spontaneous generation did not arise as the consequence of any particular philosophy. For them, spontaneous generation was simply an obvious, empirically established fact, the theoretical basis of which was of secondary importance."

The view that living organisms could develop spontaneously persisted through the Middle Ages into the 19th century. Even such advanced thinkers as the English philosopher Francis Bacon (1561–1626); the Flemish physician J. B. van Helmont (1577–1644), who introduced the concept of a gas into science; the English physician William Harvey (1578–1657), the originator of the theory of the circulation of the blood; and the French philosopher René Descartes (1596–1650) accepted spontaneous generation of some organisms as a proven fact.

In 1666, some doubt was cast on the established view that life can arise spontaneously by the Tuscan physician Francesco Redi. He showed that meat covered with muslin, to prevent access of flies, did not produce maggots, whereas such creatures did develop in unprotected meat. He concluded, correctly, that flies had laid their eggs on the uncovered meat and that the maggots had hatched from the eggs. Nevertheless, Redi

was not prepared to relinquish completely the notion of the spontaneous generation of life. He thought that, in suitable circumstances, living creatures were in fact produced directly from rotting materials.

With the invention of the microscope at the end of the 16th century, a new element was introduced into the problem of the origin of life. By means of a combination of magnifying lenses, A. van Leeuwenhoek (1632–1723), in Holland, was able to detect in air and water, and elsewhere, many hitherto invisible organisms. Such microscopic organisms (or micro-organisms) were found to occur in large numbers in decaying and rotting material, and van Leeuwenhoek argued that they did not arise spontaneously but resulted from the multiplication of organisms present in the air.

Some confirmation of this view was provided by the results of work of the French scientist L. Joblot, published in 1718, and that of the Italian Abbé L. Spallanzani, reported in 1765. They showed that if a nutrient medium, such as plant infusion, is heated to destroy the existing micro-organisms, no further organisms would develop unless the medium was exposed to the air. The experiments were criticized by contemporary scientists and belief in spontaneous generation persisted for almost another hundred years. Not only were the materialistic natural philosophers, the scientists of the day, convinced of the reality of the spontaneous generation of living organisms, but so, also, were many Christian theologians.

Although Father Spallanzani rejected the idea of spontaneous generation of life, this concept was not considered to be contrary to the teachings of the Bible. In the first chapter of Genesis it is stated, in verses 11, 20, and 24, that God commanded the Earth and the water to "bring forth" various forms of life. It did not seem unreasonable that this would be a continuing process and, consequently, it was accepted that living organisms could be "brought forth" from the proper environment. Among the eminent clerics who believed in spontaneous generation were St. Basil the Great of the Orthodox Christian Church and St. Augustine of Hippo of the Roman Church in the 4th century, and, in the 13th century, the Dominican Friar Albertus Magnus and his famous pupil St. Thomas Aquinas.

A still later supporter of the concept of the spontaneous generation of life was the English priest Father J. T. Needham who, in 1749, claimed to have observed the spontaneous production of living organisms in various liquors and infusions even when air was excluded. In 1776, Needham entered into a controversy with his fellow priest Spallanzani. He contended that the reason the latter had failed to detect the development of life was that he had heated his nutrient medium so strongly that its generative power had been destroyed. In a manner of speaking, Needham was, of course, correct, because the so-called generative power depended on pre-existing organisms and these were destroyed by heating.

Arguments for and against the view that life could arise spontaneously continued among philosophers and experimenters until the early 1860's, when the idea was dealt its death blow by Louis Pasteur in France. Even at that time there were numerous skeptics, but as a result of his many careful experiments, Pasteur refuted one objection after another. He was thus able to establish, without a shadow of doubt, that living organisms do not develop in sterile nutrient media in the absence of air or in the presence of air from which all micro-organisms have been removed. When exposed to ordinary air, however, the same media produced such organisms in large numbers.

"There is no circumstance known today," wrote Pasteur in 1862, "whereby one can

affirm that miscroscopic beings have come into the world without germs, without parents resembling themselves." Soon afterward, this view became universally accepted among scientists, and the general feeling may be summed up in the words of the famous physicist William Thomson (Lord Kelvin) who, in his presidential address to the British Association in 1871 said: "Dead matter cannot become living without coming under the influence of matter previously alive."

The Panspermia Hypothesis

With the overthrow of the theory of the spontaneous generation of life, scientists found themselves in the situation of having to choose among three possibilities, none of which was particularly attractive to them. The alternatives were either that life was created in a supernatural manner by a divine act, that life had no beginning, or that the first living organism had evolved from inorganic (nonliving) sources.

Many scientists were inclined to accept the somewhat frustrating view that life has always existed and that it was useless to consider how it originated. In 1868 the eminent German chemist Justus von Liebig wrote: "We may only assume that life is just as old and just as eternal as matter itself, and the whole controversial point about the origin of life seems to be disposed of by this simple assumption." The same view was expressed by the distinguished German scientist, physician, and philosopher H. L. F. von Helmholtz in 1874. "It appears to me," he wrote, "to be a fully correct procedure, if all our efforts fail to cause the production of organisms from nonliving matter, to raise the question of whether life has ever arisen, whether it is not just as old as matter."

The hypothesis that life had no beginning led to a revival of the concept of panspermia, meaning the widespread distribution of seeds (or life). In its original form, the idea appears in the writings of the Greek philosopher Anaxagoras (500–428 B.C.) who suggested that the life-giving principle takes the form of invisible germs which are present everywhere and are able to produce life from nonliving matter. The panspermia theory received some minor support throughout the years, but it was only after the spontaneous generation of life was shown to be highly improbable, if not impossible, that it attracted serious attention. It was supposed that the germs of life are spread throughout the universe and when they fall on a body, such as a planet, where the conditions are favorable, they will establish themselves and become the ancestors of all living organisms on that planet.

As far as life on Earth was concerned, several scientists, including William Thomson in the United Kingdom, H. von Helmholtz in Germany, and the botanist P. van Tieghem in France, thought that the germs of life might have been brought to the planet by meteorites. It is true that, in their passage through the atmosphere, meteorites become very hot on the outside, so that life germs would undoubtedly be destroyed, but the interiors of freshly fallen meteorites have been found to be quite cold. The discovery in 1834 that certain meteorites, the carbonaceous chondrites (ch. IX), contain quite complex organic (carbon) compounds seemed to provide support for the idea that meteorites might be the carriers of life germs.

Several investigators, however, including Pasteur, were unable to find in meteorites any organisms capable of multiplying in a nutrient medium. The American biologist C. B. Lipman reported in 1932 that he had obtained living bacteria from several specimens of stony meteorites, but the general opinion was that they were terrestrial organisms which had contaminated the meteorite specimens. The

319-889 O—69——11

view has been expressed in recent years that carbonaceous chondrites contain fossils of living organisms, but such organisms, if they are real, have long been dead. There is no serious suggestion that they are viable and could multiply in a nutrient medium.

An alternative form of the panspermia theory that attracted considerable interest was proposed in 1906 by the Swedish physical chemist Svante Arrhenius, who devoted much of his later life to the study of cosmical problems. He suggested that microscopic spores were ejected from a remote planet by electrical forces and that they were carried through interstellar space, from one planet to another, propelled by the pressure of radiation.

Immediate objection to this form of the panspermia hypothesis was raised on the grounds that, in their long journey through space, the spores would be exposed to ultraviolet radiation which is known to be lethal to micro-organisms. In reply, Arrhenius claimed that the bactericidal effect of ultraviolet radiation is caused by the presence of oxygen, which becomes activated and kills the organisms. Because there is no oxygen in space, the spores would survive their long journey. Experiments by Pierre Becquerel in France in 1910, however, showed that bacteria are destroyed by ultraviolet light even in the absence of oxygen.

The matter of the survival of micro-organisms or spores when traveling long distances through space, exposed to extremely low temperatures, an almost complete vacuum, as well as to various harmful radiations, was reexamined in some detail in 1961 by Carl Sagan in the United States. One of his conclusions is that a micro-organism ejected from Earth would be killed by radiation from the Sun before it reached the orbit of Neptune. This means that a similar organism coming into the solar system from the vicinity of another star, similar to the Sun, could not survive its journey to Earth. Although there is insufficient evidence for completely rejecting the panspermia theory, as far as other planets more distant than Earth from a central sun are concerned, it seems certain that spores from space could not have been responsible for life on Earth.

Development of Life From Inorganic Matter

Even if there were no other objections to the panspermia concept, and to the related idea that life had no beginning, such theories are intellectually unsatisfying because they lead to a dead end. If life spores are universal and have always existed, then life had no origin and no useful purpose would be served by pursuing the problem any further. A similar point of view, which he apparently changed later, was expressed by the famous scientist Charles Darwin in 1863. In a letter to his friend J. D. Hooker, the English botanist, he wrote: "It is mere rubbish, thinking at present of the origin of life. One might as well think of the origin of matter."

If emphasis is placed on the words "at present," Darwin's ridicule was quite justifiable at the time. During the middle years of the 20th century, however, scientists began to think seriously, and with some success, about the origin of matter. Furthermore, advances in biology and related subjects made it possible, as will be seen in due course, to develop a serious scientific approach to the question of the origin of life.

The current attitude of many biologists concerning the origin of life is based on the views which appear to have been first formulated in 1820 by the French naturalist J. B. P. Lamarck as an extension of his theory of evolution. "Among the inorganic bodies," he stated, "[there evolved] . . . extremely small, half-liquid bodies of very diffuse consistency . . . [which] developed further into

cellular bodies having an outer envelope with liquid contained in it acquiring the first rudiments of organization."

The idea that life developed from inorganic matter was accepted by the materialistic philosophers of the 19th century, such as Friederich Engels in Germany and Herbert Spencer in England. The doctrine of the evolution of biological organisms, from the simplest to the most complex, was becoming widely accepted among scientists during the 1860's, and Engels and others postulated that life itself was the result of a continuous evolution of matter. In the words of the American biologist John Keosian, the feeling was that "the origin of life . . . [is] only a rung in the long ladder of development."

The materialistic point of view concerning the origin of life was accepted by many prominent scientists in the 1870's. For example, the biologist T. H. Huxley and the physicist John Tyndall in their presidential addresses to the British Association, in 1870 and 1874, respectively, discussed the possibility that life may have originated in some manner from lifeless matter. Furthermore, in a letter to the English naturalist Alfred R. Wallace, written in 1872, Charles Darwin stated: "On the whole it seems to me probable that Archebiosis [i.e., the theory that life developed in the remote past from nonliving materials in a series of steps] is true."

In answer to the objection that if life had come from nonliving matter in the past, it should still be doing so today, Darwin wrote as follows:

> It is often stated that all the conditions for the first production of a living organism are present which ever could have been present. But if . . . we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc., present, that a proteine compound was chemically formed ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed.

As a matter of fact, the atmospheric conditions when "the first production of a living organism occurred" were probably very different from those existing at present. Moreover, in 1961, J. Keosian argued that it is indeed possible that life is still developing from nonliving matter. Nevertheless, the main point of present interest is that Darwin thought a living organism might have been produced from inorganic substances "in some warm little pond" under the influence of "light, heat, and electricity." This is not substantially different from the view widely held by biologists today.

Speculations concerning the manner in which life could have arisen from nonliving substances present in nature continued throughout the later years of the 19th century and during the early part of the present century. Two publications in the 1920's, however, stand out as marking the beginning of a new era in the evolution of ideas concerning the origin of life on Earth.

The first of these publications, by A. I. Oparin, was called "The Origin of Life." It appeared in 1924 but was based on a lecture delivered 2 years before. Oparin developed in some detail the materialistic argument of the gradual step-by-step evolution of life from the simplest compounds of the common elements, hydrogen, carbon, nitrogen, and oxygen, through more complex substances to primitive living organisms.

A somewhat similar point of view was expressed independently in 1928 by the English biochemist J. B. S. Haldane. He suggested that various organic compounds [1] were formed

[1] The term "organic compounds" is used here in its general sense to include all carbon compounds, regardless of whether they are produced by living organisms or not.

as a result of the action of ultraviolet light on the gases of the atmosphere and that "before the origin of life they [i.e., the organic compounds] must have accumulated till the primitive oceans reached the consistency of hot dilute soup." In this "soup," which contained the basic chemicals of life, Haldane thought that life itself began. It is of interest that the word "soup," now commonly used as a scientific slang term for the medium in which life first evolved, appears to have had its origin in Haldane's statement just quoted.

The speculations of Oparin and of Haldane should have led to a considerable experimental program to determine if prebiological (or prebiotic) compounds, the organic chemical compounds from which life might have evolved, could be produced under simulated primitive atmospheric conditions by the action of "light, heat, electricity, etc.," as suggested by Darwin. Such work as was done in this area, however, did not appear to give promising results, largely because the notions concerning the composition of Earth's original atmosphere were probably wrong.

The views expressed by Oparin, in an extended elaboration of his original thesis entitled "The Origin of Life on Earth," published in 1936, are in general agreement with those now held. Nevertheless, it was not until after 1952, when the American scientist H. C. Urey reported the results of his thermodynamic calculations on the composition of Earth's primitive atmosphere, that the situation took a new turn. The consequences of this development will be described in chapter IX.

Another possible reason the question of how life originated did not attract great interest, at least not among experimentalists, is that a unifying principle was missing. Such a principle became apparent, however, in the 1950's with the growth of the field of molecular biology, which resulted in a better understanding of the basic phenomena of life in terms of molecular structure.

What Is Life?

Before proceeding to describe recent developments in molecular biology, it is opportune to consider the fundamental question, which has been avoided so far: What is life? If the evolutionary view is accepted, that living organisms developed from inorganic matter as the result of a number of gradual changes, then it may be impossible to draw a line of demarcation between that which is not and that which is alive.

In 1937, the British biologist N. W. Pirie argued that an exact definition of life is "meaningless" and went on to say that "it seems prudent to avoid use of the word 'life' in any discussion about border-line systems and to refrain from saying that observations on a system have proved that it is or is not alive." Other writers have accepted this point of view and have concluded, in the words of J. Keosian, that "attempts at an exact definition of life are not only fruitless, at least at present, but meaningless."

Even if an exact definition of life, which will cover borderline (or transitional) cases, cannot be formulated, it is, nevertheless, useful to try to describe the essential characteristics of such systems as are indisputably alive. It is not meant to imply, however, that systems not possessing these characteristics do not have "life." Until some indefinite time in the future, a time which many molecular biologists think is not too far distant, when the transitional systems are produced in the laboratory, a rigorous definition of "life" will not be possible—and perhaps not even then! Meanwhile, however, an understanding of the fundamental aspects of life, as it is known at present, will provide a focal point from which to consider the chain of evolutionary

processes that, it is assumed, led ultimately to a living organism.

One essential characteristic of an organism which is definitely alive is the property of replication: the capability of producing a copy (or replica) of itself. In the simplest micro-organisms, the organism as a whole replicates, but in the more complex forms of life, the replication occurs among the individual cells which make up the organism. With this limitation in mind, it can be stated that a living system can select suitable materials from its environment and can convert them by a complex series of processes into a copy of itself.

As a general rule, the copy is an exact one, but occasionally there are mistakes, and to describe a living system on the basis of replication only is insufficient. In an autocatalytic chemical reaction, for example, one of the products of the reaction stimulates (catalyzes) its own formation. The rate of increase in the quantity of this product is represented by the same type of mathematical expression as that for the rate of increase, or growth, of a simple micro-organism in a nutrient medium. A flame provides another instance of replication in a system that is not regarded as possessing life.

It is evident, therefore, that a further qualification is required to distinguish living from nonliving systems. Such a qualification arises from the cases, which occur from time to time, when there is an error in copying. These occasional accidental variations in replication are called mutations, or mutants. The important characteristic of living systems is that the mutants copy themselves, and not the original (parental) forms, when they replicate. Some of the mutants are less suited than were the parental forms to cope with their environment. Such mutations, which are said to be deleterious, will soon die out. Other mutations, on the other hand, are beneficial, and the mutant forms will persist and multiply, often at the expense of the unmutated (or parental) type. It is the occurrence of mutations, which copy themselves and not the parental form, that has made organic evolution possible.

A definition of life based on the foregoing ideas was proposed by the American geneticist H. J. Muller, and it is well expressed in the words of N. H. Horowitz in the United States in 1966:

> The most concise, unambiguous general definition of life that can be given at the present time is based on the genetic properties of living things. According to this view, the unique attribute of living matter from which all of its remarkable features derive is its capacity for self-duplication with mutation. That is to say, living organisms are systems that reproduce, mutate, and then reproduce their mutations.

For the present purpose, therefore, it will be postulated that if an organism, or system, is capable of replication and mutation, with replication of the mutant, then it is definitely living.

Cells in Living Organisms

All known living systems are made up of cells and it is in the cell where the processes leading to replication and mutation occur. The very simplest organisms, such as the ameba and other protozoa and many bacteria, consist of a single cell, whereas more complex organisms, such as man, contain thousands of billions of cells. With but a few exceptions, all cells have similar basic structures, although their functions may differ widely. The striking resemblances between the cells of living organisms of all types suggest a common origin of all forms of life.

Although a great deal more is now known about cells, a statement made in 1836 by the German biologist Theodor Schwann, one of the founders of the cell theory of the structure of living systems, still serves to summarize the

situation. "All organisms," wrote Schwann, "are composed of essentially like parts, namely cells. These cells are formed and grow in accordance with essentially the same laws; hence . . . the processes must everywhere result from the operation of the same forces." It would not be unreasonable to conclude, therefore, that all living organisms have developed from the same ancestor, or ancestral type. There is other evidence for this view.

Most cells have dimensions which are measured in terms of a few microns (p. 76), some ten-thousandths of an inch. They are consequently visible only in a microscope. In spite of their small size, cells usually have complex structures, but for the present purpose it is sufficient to consider only some of the basic features. These are indicated in figure 8.1 and are fairly typical of all except some primitive single-celled organisms, such as the bacteria and the blue-green algae. The cell wall (plasma membrane) encloses a viscous liquid material, called the cytoplasm, containing various bodies of which the most conspicuous is the nucleus. The latter is surrounded by a somewhat porous membrane and within it are a more-or-less spherical nucleolus and some irregular masses of darkish material known as chromatin. A number of small bodies, referred to as ribosomes, are also to be found in the cytoplasm. They are visible in the microscope only under very high magnification.

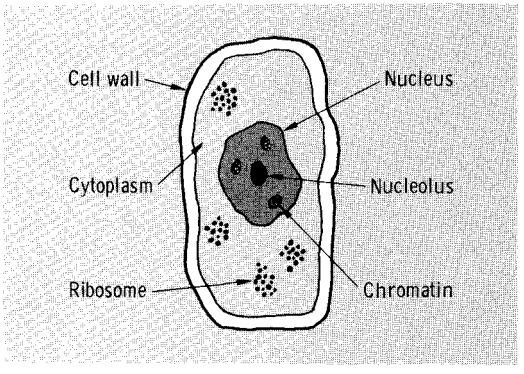


FIGURE 8.1. Basic features of a living cell.

When a cell reaches a certain stage of development, it divides and produces a replica of itself. In this process of mitosis, as it is called, the chromatin within the nucleus is seen to become organized into a number of threadlike structures, known as the chromosomes, and the nucleolus seems to disappear. The genetic properties of the organism, that is, the capacity for replication and mutation, are associated with the chromosomes. The number of chromosomes in a nucleus depends on the particular species. In man, the nuclei of cells other than those involved in reproduction contain 46 chromosomes.

In the course of mitosis, the chromosomes form into two groups, each of which is an exact replica of the other. This is the essential basis of replication. The two groups separate and move to opposite ends of the cell (fig. 8.2), where they tend to become less visible. Within a short time, a new nucleolus and a surrounding nucleus form at the ends of the original cell. The latter then divides into two daughter cells, each containing its own cytoplasm, nucleus, and other components just like the parent cell. The two cells formed in this manner can then undergo mitosis to produce four cells and so on.[2]

The chromosomes are divided into a number of sections, several thousand in a complex organism like man. These are called genes and are thought to carry the individual hereditary, or genetic, characteristics of the

[2] The bacteria and blue-green algae do not have a definite nucleus surrounded by a distinct membrane. The whole cell is then roughly equivalent to a nucleus and replication takes place by the process called fission. Apart from this difference, the genetic behavior is believed to be the same as described below for cells containing nuclei.

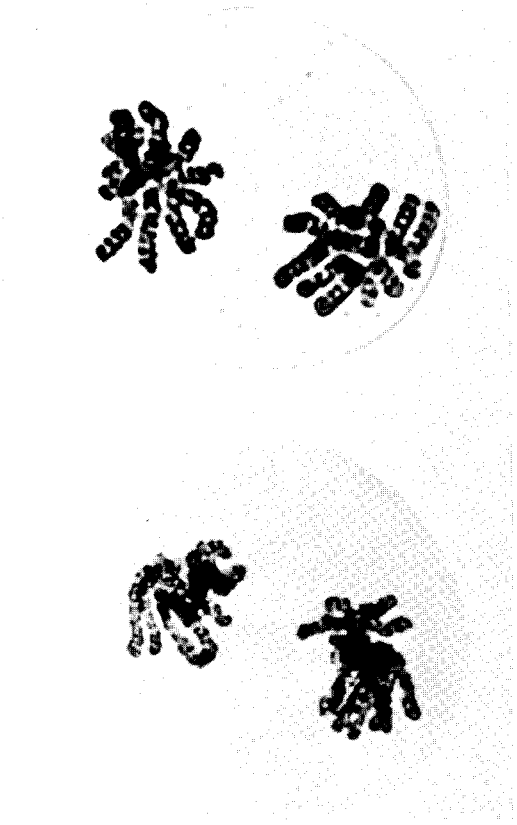


FIGURE 8.2. Early stage of mitosis showing separation of two groups of chromosomes. (Brookhaven National Laboratory photograph.)

organism. In normal circumstances, the chromosomes in each daughter formed by mitosis have the same genes as in the parent cell. In this way, the hereditary characteristics of the cell or organism are continued when the cell replicates. Mutations are ascribed to changes in the arrangement or nature of the genes within a chromosome. When a mutant cell replicates during mitosis, it produces a chromosome just like itself, and not like the original unmutated chromosome. Hence, once a mutation occurs in a cell, the organism continues to produce mutant forms of the cell.[3]

In addition to their ability to replicate, and thereby perpetuate genetic characteristics of the cell or organism, the chromosomes have another important function. The individual genes apparently carry the information which is translated into instructions for the cell to produce the great variety of proteins, including the enzymes which act as catalysts for the many chemical processes occurring within the cytoplasm.

MOLECULAR BIOLOGY AND THE ORIGIN OF LIFE

Proteins

One of the outstanding achievements of biology in recent years has been the identification of the functions of the chromosomes with the nature and properties of certain specific molecules. The results of studies in molecular biology have thrown much light on the mechanisms of replication and mutation, and on the production of proteins. The conclusions appear to have an important bearing on the problem of the origin of life on Earth. Consequently, an outline of some of the basic aspects of molecular biology will be given here.

Two types of molecules, proteins and nucleic acids, play fundamental roles in the functioning of a living cell. The important characteristics of the chemical structure and nature of these two classes of substances will be considered in turn.

A protein is a large molecule consisting essentially of a number of amino acids joined

[3] In organisms which reproduce sexually, the mutation must occur in a germ cell, either the ovum in the female or sperm in the male, if it is to be carried over into the next generation. Mutations in other (somatic) cells will replicate and multiply within the body of the organism, but they will not be passed on to subsequent generations.

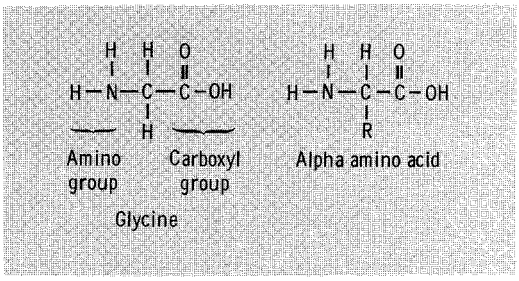
together to form a long chain. An amino acid is a carbon compound containing an acidic (—COOH) group and also a basic amino ($—NH_2$) group. Some 20 different amino acids have been found to occur in proteins, although a given protein does not necessarily contain all 20 amino acids.[4] In any event, the proportions of the different amino acids may vary widely from one protein to another. Furthermore, there are a very large number of different possible sequences in which the amino acids can be connected to form the protein chain. The characteristic properties of a particular protein are determined by the nature of the amino acids it contains, their relative proportions, and the order in which they appear in the chain. A change in any one of these three highly specific attributes will change the characteristics of the protein molecule.

Because a protein chain consists of a large number of amino acid molecule residues, the total number of possible proteins is fantastically large. Consider, for example, insulin, which plays an important role in controlling the sugar content of the blood. Although insulin is one of the simplest known proteins it contains a chain of 51 residues of 17 different amino acids. It has been calculated that these could be arranged in more than 10^{50} different ways, yet only one of this enormous number of possibilities represents the insulin produced in the human pancreas. There are, however, a few other forms, different in minor respects from human insulin, that serve the same purpose in animals.

Proteins have been compared to words made up from an alphabet consisting of 20 different letters. Each protein "word" may contain from about 50 to several hundred letters, selected in any numbers and arranged in any order from the 20 available. Obviously the total number of possible words will be extremely large. Most of the words, however, have no meaning for a given organism. These would represent proteins that either do not exist in nature or are not produced by that organism. Each "intelligible" word has a specific meaning that is different from that of other such words. That is to say, each protein has a definite and specific function in the organism that produces it.

The proteins of particular interest for the present are those which act as enzymes, that is, as catalysts, for the numerous processes occurring in a living system. Each of these processes requires a specific catalyst, represented by a protein word with the appropriate meaning. Some enzymes embody additional atomic groupings, but the protein is always the main component of the molecule. Living organisms also contain many other proteins which occur in body tissues but are not enzymes. These are of secondary interest only, however, for the present problem, the origin of life.

The simplest amino acid found in proteins is glycine or aminoacetic acid, and the other 19 may be regarded as being derived from, or related to, glycine. The structural formula of glycine (NH_2CH_2COOH) is shown at the left below:

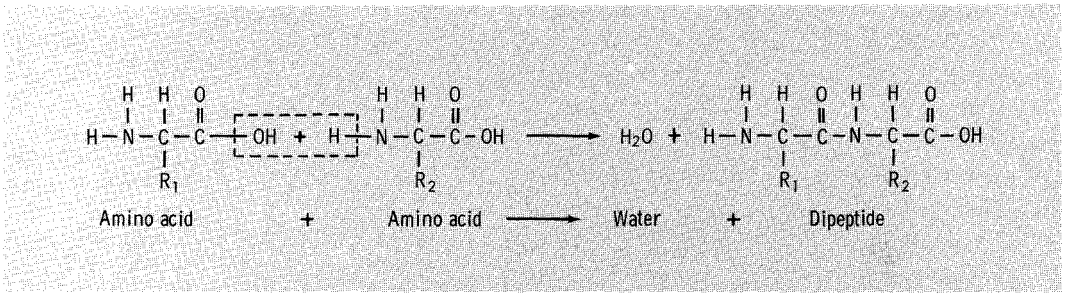


[4] Slightly different numbers are often given in the literature. The exact number depends upon whether certain pairs of related amino acids are considered as one or two individuals. The choice is somewhat arbitrary, but the number 20 is in common use.

All the amino acids in proteins may be represented by the general formula at the right of the glycine formula. It will be noted that, in the latter, one of the hydrogen (H) atoms of the CH_2 group, which links the amino and carboxyl groups in glycine, has been replaced by R. The letter "R" represents what is called a side chain of the amino acid. In glycine itself, R is merely a hydrogen atom, but in the other amino acids, R ranges from a simple group, like $-CH_3$ (in alanine), to carbon and carbon-nitrogen rings in phenylalanine and proline, respectively. The details of these side chains are, however, not of immediate concern.

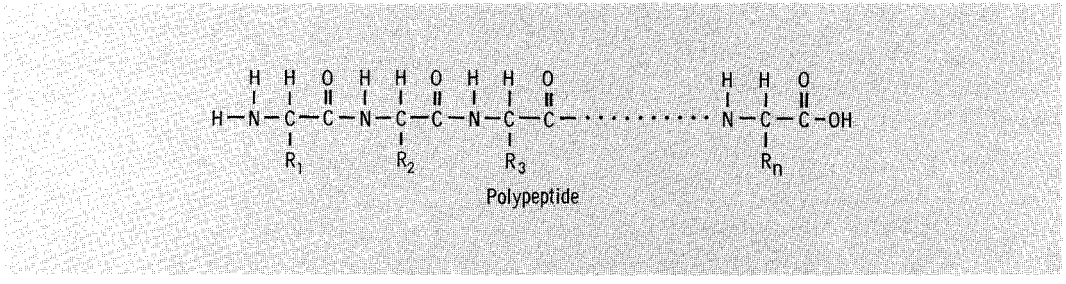
An amino acid in which the amino and carboxyl groups are attached to the same carbon atom, as they are in glycine, is called an alpha amino acid. Because all the amino acids found in proteins are derivatives of glycine, they are all alpha amino acids. This fact may have some bearing on the origin of living organisms.

The amino group has the properties of a base, whereas the carboxyl group confers acidic properties on a molecule. The amino group of one amino-acid molecule can interact with the carboxyl group of another molecule, which may be of the same or different type:



where R_1 and R_2 may be the same or different side chains. A molecule of water (H_2O) is split out and the product is sometimes called a conjugated amino acid because it is formed by the conjugation of two simple amino acids.

The conjugated amino acid, like all amino acids, has an amino group (at the left) and a carboxyl group (at the right). Consequently, it also can conjugate with another amino acid, which may be simple or conjugate. Thus, it is possible to build up a chain of amino acids, and the structure of such a chain may be generally represented by the formulation shown below:



where n is the total number of conjugated amino-acid residues in the chain. The side chains, R_1, R_2, R_3, . . . , R_n, may be the same or different.

It will be noted that the two members of each pair of amino acids in a chain are joined by a —CO·NH— linkage; this is called a peptide bond. A molecule containing only a few amino acids joined together in a chain is often referred to as a peptide (or peptide molecule).[5] If there are only two amino-acid residues in the molecule, the name dipeptide is used. Three amino-acid residues give a tripeptide, and so on. When the number of amino-acid residues is fairly large, the molecule is referred to as a polypeptide. A protein is a polypeptide, usually a complex one, that is found in nature.[6]

Proteins and polypeptides in general are sometimes called polymers, which is the name given to molecules built up from a large number of residues of simple molecules, known as monomers. In polypeptides the monomers are the amino acids. The term macromolecule, meaning large molecule, is also used to describe proteins, polypeptides, and other polymers.

In the formula given above to represent a polypeptide, the —N—C—C—N—C—C—N— chain is shown as forming a long straight line. In fact, however, the chain has a complex zigzag structure, with the side chains R_1, R_2, etc., alternating on one side and the other side of the zigzag backbone. The latter sometimes takes the form of a helix, a three-dimensional spiral, like a spiral staircase, which may be folded into a compact form.

In some amino acids, such as cysteine and methionine, the side chain contains a sulfur atom. The sulfur atom in one such amino acid, particularly cysteine, can link up with a sulfur atom in another amino acid. The two sulfur atoms may be in the same polypeptide chain or they may be in different chains. In the latter case, the two chains are connected together by an —S—S— bond, or bonds. In insulin, for example, there are two polypeptide chains and three —S—S— bonds. One of these three bonds is between two cysteine residues in the same chain, and the others between such residues in different chains.

Optical Activity of Amino Acids

An examination of the general formula of alpha amino acids given earlier will show that the alpha carbon atom, that is, the one to which both —NH_2 and —COOH groups are attached, is linked to four different atoms or groups of atoms: NH_2, H, R, and COOH. In glycine, R is H, so the four groups are not all different. Hence the following discussion does not apply to glycine, but it is applicable to all the other amino acids found in proteins.

When a molecule contains a carbon atom to which four different atoms or groups are attached, the carbon atom is said to be asymmetric. Every molecule with such an asymmetric carbon atom can exist in two distinct structural forms, each being a mirror image of the other. The two configurations have been compared to a pair of gloves. The gloves are identical in structure yet each differs from the other. An important difference between the two structural forms of a molecule containing an asymmetric carbon is observed when plane polarized light (p. 85) is passed through a quantity of each in either the pure state or a solution. One form rotates the plane of polarization to the right (dextrorotatory form), whereas the other rotates the plane of polarization by the same angle to the

[5] The term oligopeptide is sometimes used, from the Greek *oligos,* meaning a few.

[6] According to one convention, peptides of molecular weight up to 10 000 are called polypeptides and above that they are referred to as proteins.

left (levorotatory form). The two forms are called optical isomers and the molecule, or compound, is said to possess optical activity.

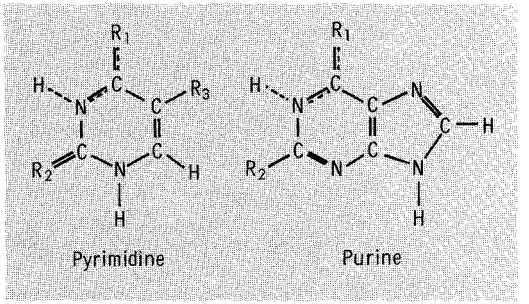
With the exception of glycine, all the amino acids in proteins contain at least one asymmetric carbon atom, so they are optically active. In each case, there are at least two spatial configurations, designated by the letters D and L, respectively. The D-forms all have a configuration related to the dextro form of the amino acid serine, whereas the L-forms are related to the levo form of this amino acid. Although D-serine rotates the plane of polarized light to the right (dextrorotatory) and L-serine to the left (levorotatory), this is not necessarily so for all amino acids or polypeptides. Thus, the designations D and L refer strictly to types of configuration or arrangements in space of the four different atoms or groups attached to the alpha carbon atom in the amino acid.

If amino acids are synthesized in the laboratory, as they all can be, the product consists of a mixture of exactly equal proportions of the D- and L-forms. This is called a racemic mixture of the optical isomers and it does not rotate the plane of polarization of light. It is a striking fact, however, that, with but a very few minor exceptions, the amino acids found in the proteins of all living organisms, from ameba to man, always have the L-configuration. How and why this preference for the L-forms of amino acid has arisen are among the major unsolved problems of biology.

Nucleic Acids

The nucleic acids, as their name implies, are formed in the nucleus, but some are produced in other parts of the cell. Like the proteins, the nucleic acids are macromolecules, or polymers, which may be regarded as composed of chains of simpler units. The fundamental unit of structure of nucleic acids is itself a fairly complex molecule called a nucleotide. A nucleotide consists of three distinct subunits: residues of a nitrogenous base, a sugar with five carbon atoms, and phosphoric acid.

The nitrogenous base belongs to one or the other of two related groups. These are pyrimidines and purines, of which the general structures are indicated below. The symbols



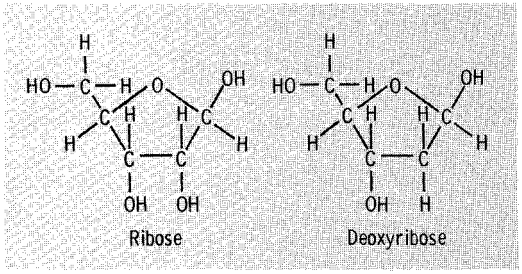
R_1, R_2, R_3 may be either hydrogen or oxygen atoms or various groups, such as CH_3 or NH_2. It will be noted that the left-hand ring of atoms in the purines is identical with the pyrimidine ring. Of the nitrogenous bases which are found in nucleic acids, three pyrimidines—cytosine, thymine, and uracil—and two pyridines—adenine and guanine—are by far the most common. The identities of R_1, R_2, and R_3 in these five bases are given in the accompanying table.[7]

	R_1	R_2	R_3
Pyrimidines:			
Cytosine	$-NH_2$	$=O$	$-H$
Thymine	$=O$	$=O$	$-CH_3$
Uracil	$=O$	$=O$	$-H$

	R_1	R_2
Purines:		
Adenine	$-NH_2$	$-H$
Guanine	$=O$	$-NH_2$

[7] When the group R_1 is NH_2, the H atom attached to the N atom at the upper left is absent.

The sugar component of the nucleotides is a five-carbon sugar. This is either ribose or the related deoxyribose in which one of the —OH groups of ribose is replaced by a hydrogen atom:

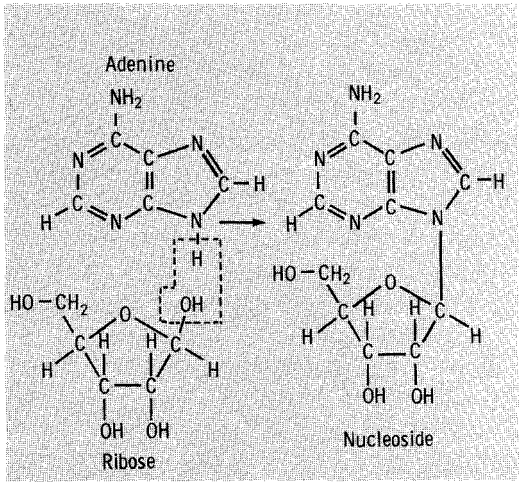


Consequently, deoxyribose has one oxygen atom less than ribose. Nucleotides containing deoxyribose are sometimes called deoxynucleotides where a distinction is desirable. In general, however, the term nucleotides includes compounds of the same type regardless of whether the sugar is ribose or deoxyribose.

In the nucleotide, the sugar residue is attached to the nitrogenous base by elimination of a molecule of water. The hydroxyl (—OH) group indicated at the right of the sugar structure combines with the hydrogen atom of the amino (=NH) group of the pyrimidine or purine shown at the bottom of the formulas depicted on page 165. Thus, the reaction of ribose with the purine base adenine may be represented by the structural formula shown at the top of the next column.

The combination of nitrogenous base with a five-carbon sugar is called a nucleoside. The nucleoside formed from a pyrimidine base is quite similar. The sugar is attached in the same way to the nitrogen atom shown at the bottom of the structural formula.

It will be observed that the four carbon atoms in the ribose ring are all asymmetric, as also are three of those in the deoxyribose ring, so these sugars exist in optically active forms.

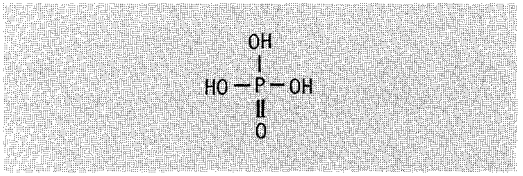


The particular configurations of both ribose and deoxyribose found in nucleic acids are always the D-modification. In fact, all sugars (or sugarlike) compounds produced or utilized by living organisms are of the D-form, and their spatial configurations are related to that of the dextrorotatory modification of glyceraldehyde.

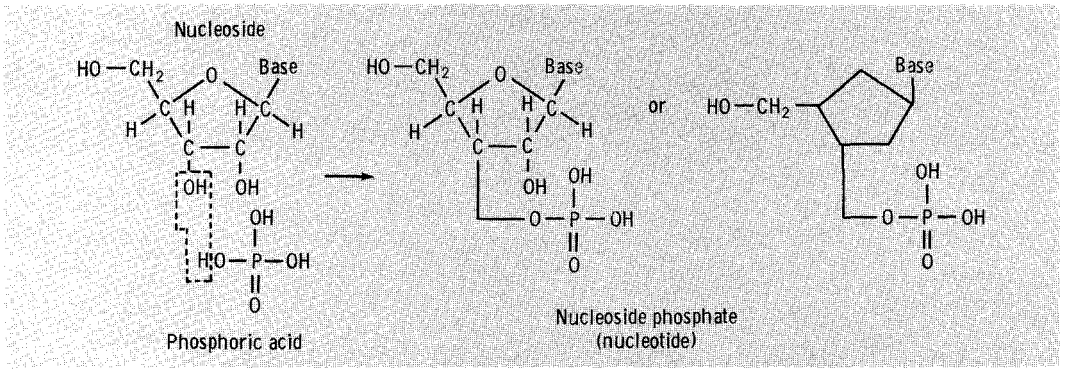
It should be pointed out that the preference in nature for the D-forms of sugars as against the L-forms of amino acids has no significance. The choice of the symbols D and L to represent the structures of the reference compounds, serine for the amino acids and glyceraldehyde for the sugars, is somewhat arbitrary. As indicated earlier, L-amino acids are not necessarily levorotatory compounds; some may rotate the plane of polarization of light to the right. Similarly, D-sugars are not always dextrorotatory. An outstanding example in this connection is D-fructose (fruit sugar), one of the two simpler components of cane sugar. It is, however, levorotatory; that is, it rotates the plane of polarization of light to the left.

The final member of the nucleotide molecule is the residue of orthophosphoric acid.

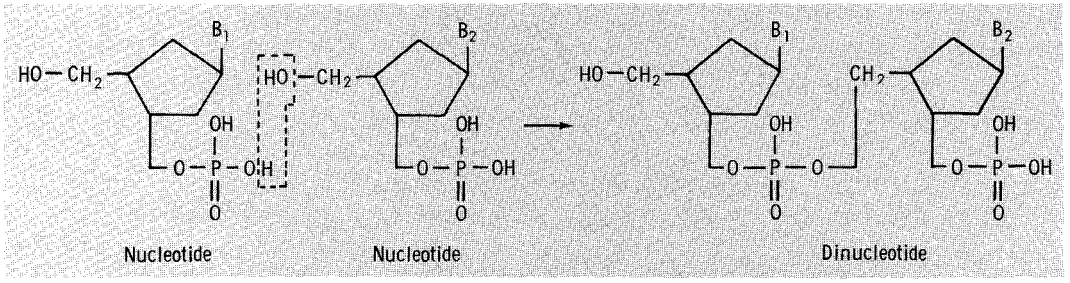
The acid itself is indicated by the structural formula



In the formation of a nucleotide, the phosphoric acid (or phosphate) residue becomes attached to the lower left carbon atom of the sugar residue in the nucleoside. The process may be represented in the following manner, where a pentagon implies a ribose (or deoxyribose) residue:

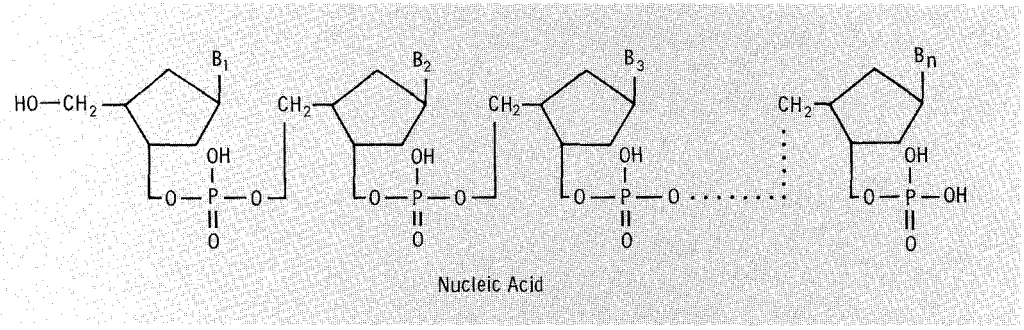


Two nucleotide molecules, which may be the same or different, can join together in much the same way as two amino acids can. A molecule of water is eliminated between the phosphate residue of one nucleotide molecule and the sugar residue in the other, in the manner indicated below.[8] If the sugar is deoxyribose, then the —OH group attached to the carbon atom indicated at the lower right of the molecule is replaced by a hydrogen atom. The nitrogenous base residues are represented by B_1 and B_2, which may be the same or different.



[8] In some nucleotides (p. 190) the phosphate residue is attached to the carbon atom of the CH_2 group shown at the upper left of the sugar formula, but in the dinucleotide, the phosphate residue is always connected between this carbon atom of one sugar residue to the lower left carbon atom of the other sugar residue, as shown above. The same is true for polynucleotides.

The combination of two nucleotides is called a dinucleotide, and this can add on another nucleotide, or dinucleotide. If the process is imagined to continue with more nucleotides, the product is a polynucleotide. A nucleic acid is a polynucleotide which may consist of hundreds or even thousands of nucleotide residues. Thus, the general formula of a nucleic acid consisting of n nucleotides may be written in the following manner:



As a rule, a given nucleic acid contains only four different bases, two of which are always pyrimidine residue and two purine residues. Thus a nucleic acid may be described as a long chain, or backbone, of alternating residues of a five-carbon sugar and phosphate, with side chains of nitrogenous bases attached to the sugar residues. Incidentally, as with the polypeptides, the backbone chain of the nucleic acids is not linear, but is twisted, often in the form of a spiral.

In a given nucleic acid or, rather, type of nucleic acid, all the sugar residues are the same. They are always either all ribose or all deoxyribose, but never both. When the sugar residues are all ribose, the nucleic acids are called ribonucleic acids (RNA). Similarly, in deoxyribonucleic acids (DNA) the sugar residues are all deoxyribose. All natural forms of DNA and RNA contain only two pyrimidine and two purine bases. There are so many nucleotide units in the chain, which can be connected with the bases in any sequence, that the total number of possible different forms of DNA and RNA is extremely large. Nevertheless, all forms of DNA have generally similar properties and functions, as also do the various types of RNA.

Deoxyribonucleic Acid

Both DNA and RNA are produced in the nuclei of nearly all living cells, but whereas the RNA leaves the nucleus and enters the surrounding cytoplasm, essentially all the DNA remains within the nucleus. The chromosomes consist basically of one or more chains of DNA, together with protein. According to current views, the genes may be identified with sections of these DNA chains. When a cell divides and produces two daughter cells, the nucleus of each of the daughters contains the same amount of DNA as did the original parent cell before it was ready to divide. Consequently, during the process of mitosis, or fission, the amount of DNA in the nucleus must double, half going into one set of chromosomes and half into the other.

The currently accepted explanation of the replication of DNA is based on the ideas put

forward jointly in 1953 by the American J. D. Watson and the Englishman F. H. C. Crick working at Cambridge University in England. According to these scientists, most, but not all, forms of DNA are made up of a combination of two parallel deoxyribose-phosphate chains twisted in such a way as to form an interlocking helix. The two chains are held together by bonds between each nitrogenous base, either a pyrimidine or a purine, in one chain and a complementary base, either a purine or a pyrimidine, respectively, in the other chain. The DNA may thus be visualized as resembling a spiral staircase, with each "step" consisting of one pyrimidine residue and one purine residue bound together (fig. 8.3.)

The double-chain DNA molecule may be thought of, more simply, as a long ladder. Both side pieces of the ladder consist of a succession of deoxyribose (sugar) and phosphate residues, and each rung is made up of a pair of nitrogenous bases. The ladder is then twisted into the form of a helix in most kinds of DNA. For the present purpose, the helical form of DNA is not important, and the molecule will be regarded as having a ladderlike structure.

Four bases, and four bases only, are found in DNA, no matter in what kind of cell it is formed. Two of these bases are the pyrimidines thymine and cytosine, and two are the purines adenine and guanine (p. 165). The rungs of the DNA ladder can be formed either by the combination of thymine with adenine or of cytosine with guanine, as shown on page 170, and in no other way. Thus, thymine attached to one sugar-phosphate side piece of the ladder is always complementary to adenine attached to the other; similarly, cytosine is always complementary to guanine. It is seen that the pairs of bases, one pyrimidine and one purine, are connected by what are

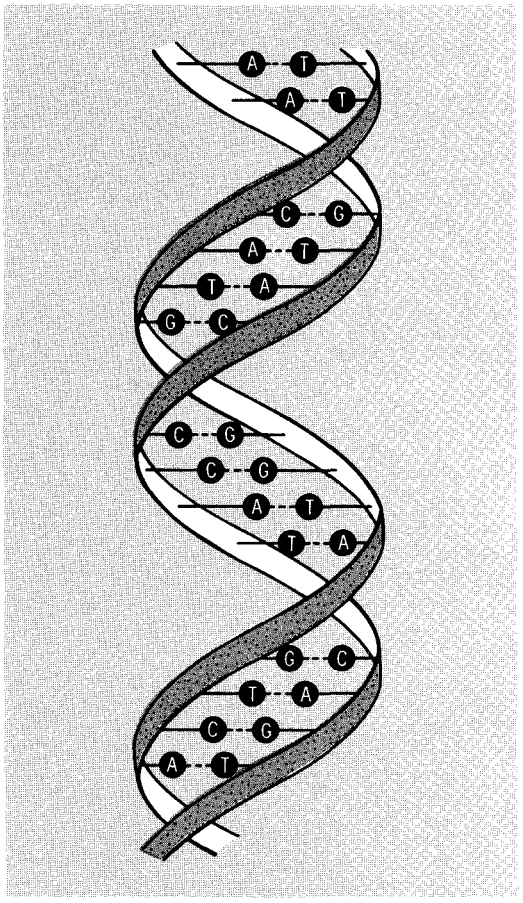
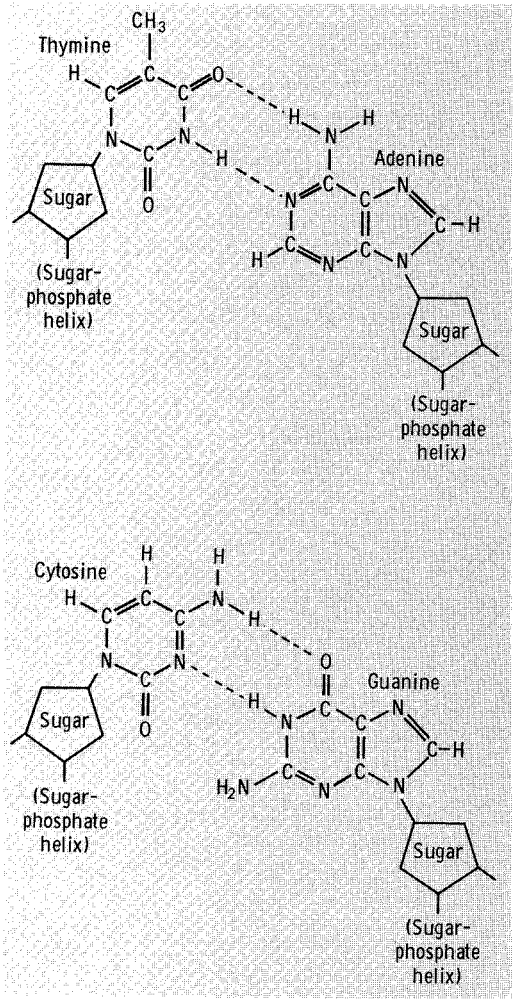


FIGURE 8.3. Representation of twisted chains in a DNA molecule. (After J. D. Watson and F. H. C. Crick.)

called hydrogen bonds. This is a relatively weak type of chemical linkage in which a hydrogen atom, originally attached to a nitrogen atom, is shared between two atoms, either two nitrogen atoms or a nitrogen atom and an oxygen atom.

For the DNA ladder to be stable, it is reasonable to expect that the rungs should be of equal length. This explains why only the combinations of thymine (T)-adenine (A)



and cytosine (C)-guanine (G) occur. The linear dimensions, as given in the structural representation above, are 11.1 and 10.8 Å, respectively. Other combinations of bases are possible in principle, but they are ruled out because they have the wrong dimensions and would cause distortion of the rungs or the side pieces of the ladder.

In the process of replication, energy is supplied to the DNA molecule, and as a result the hydrogen bonds between the nitrogenous bases are broken. The two chains then partially separate, as shown in part I of figure 8.4. The separate parts of the chain are indicated at II. The liquid in the nucleus contains the nucleotides of the four bases, and when a nucleotide with the appropriate complementary base approaches a base attached to one of the chains, combination by hydrogen bonding can occur, as shown in part III of figure 8.4. When this process is completed, as indicated at IV, it is evident that there are now two double DNA chains, each identical with the original DNA molecule. The result is that two DNA molecules have been formed, each of which is an exact replica of the original molecule.

Because of the restrictions in the manner in which pairs of bases combine, T only with A (or A with T) and C only with G (or G with C), the replication is almost invariably exact. Occasionally, however, an error may occur. For example, suppose that at one particular point in the chain the base G combines with T instead of with C, as it should. There would be some distortion in the DNA molecule at this point, but the rapid completion of the double chain above and below the erroneous rung in the ladder might permit this rung to remain.

Suppose this error occurred at the first G on the left side of the DNA in figure 8.4. The resulting molecule of DNA would then be as indicated in figure 8.5. Two different DNA molecules are now formed in the replication process. The one at the left is identical with the original DNA, but the one on the right, which satisfies all the requirements of a stable DNA molecule, is different. The latter then represents a mutation. Assuming, as is very probable, that another error does not occur immediately, this mutation will replicate itself exactly. The Watson-Crick model thus provides a mechanism for both replication and mutation and for replication of the mutant

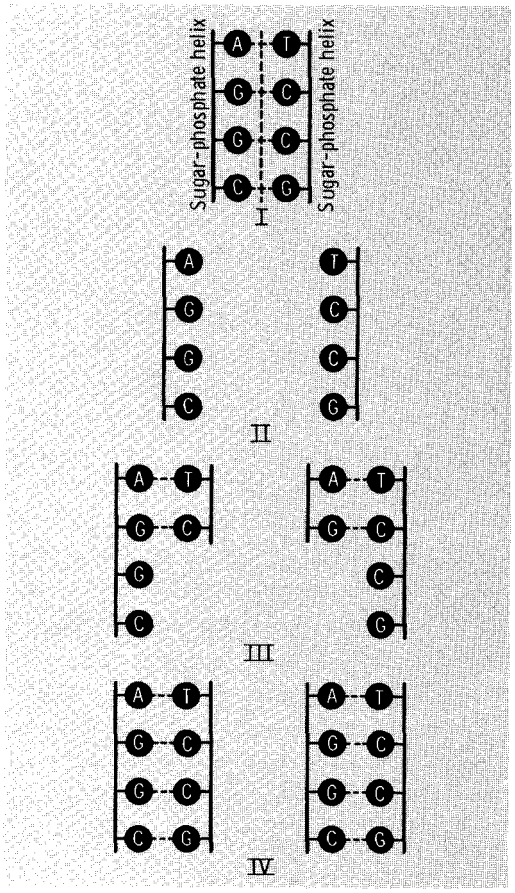


FIGURE 8.4. Schematic representation of replication of part of a DNA molecule.

DNA. There are other ways in which mutation can occur; for example, by breakage of a DNA chain followed by the parts rejoining in a different manner. In every case, however, exact replication of the mutant may be expected.

In addition to providing the essential means of replication and mutation, the DNA serves as a pattern or template for the production of one type of RNA, and this, as will be explained shortly, directs the formation of various proteins. The particular proteins produced, including enzymes, are determined by the order of the bases in the RNA, and hence by the DNA in the chromosomes. A mutation in the DNA might thus result in the generation of a new enzyme which will alter the characteristic properties of the cell. In the case of an organism which reproduces sexually, the change in the DNA must take place in a germ cell if it is to be passed on to the next generation.

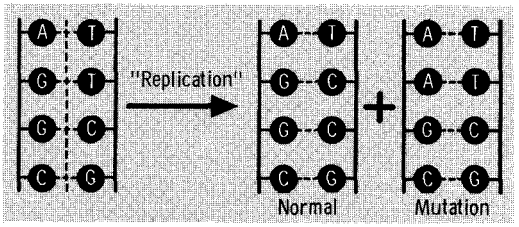


FIGURE 8.5. Possible mode of formation of a mutation in a DNA molecule.

The Ribonucleic Acids

The structure of the RNA molecule differs from that of DNA in the obvious respect that the sugar residues are ribose instead of deoxyribose. There is, however, another difference. Although there are only two pyrimidine bases and two purine bases in RNA, as in DNA, only three—adenine, cytosine, and guanine—are the same in both types of nucleic acids. The fourth base in RNA is the pyrimidine uracil (U) in place of thymine. As can be seen from the tabulation on page 165, uracil differs from thymine only in the respect that a hydrogen atom in the former has taken the place of a CH_3 group in the latter. A few other nitrogenous bases, which are closely related structurally to those mentioned above, sometimes occur in RNA molecules, but this does not have great significance.

Although it is not inconceivable for RNA molecules to exist as double chains capable of

replicating themselves, it appears that this process occurs very rarely, if at all. In general, the RNA which directs the formation of protein is produced in the nucleus by the matching of the complementary bases from nucleotides containing ribose with the bases in the existing DNA chain, as represented in figure 8.6. In this matching, U in the RNA is complementary to A in the DNA chain. A single separate DNA chain can evidently produce several different RNA molecules, or perhaps different sections in the RNA chain, each of which may be regarded as corresponding to a particular gene.

The many different RNA molecules (or sections) formed in this manner, each carrying exact specifications for a particular protein, are called messenger RNA or mRNA. The mRNA molecules, each of which is a chain of hundreds or possibly thousands of nucleotide residues, leave the nucleus and enter the cytoplasm. Here they attach themselves to the ribosome particles (fig. 8.1). The ribosomes hold the mRNA molecules as each of the latter directs the synthesis of a particular protein. The performance of this synthesis requires the aid of another type of RNA, known as transfer RNA or tRNA, which is formed in the cytoplasm. The molecules of tRNA consist of roughly a hundred or so nucleotide residues, and so they are much smaller than the mRNA molecules.

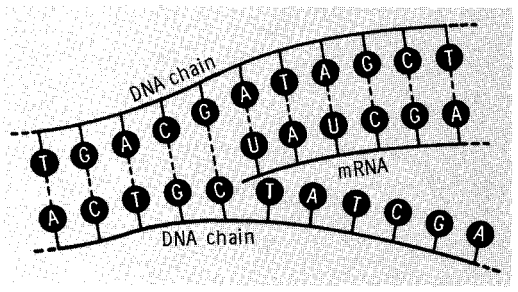


FIGURE 8.6. Production of messenger RNA.

In its chain of nucleotides, each molecule of tRNA has a group of three units which are effective in protein synthesis; the nature and order of the nitrogenous bases in this group of three nucleotide residues determine the characteristics of the particular tRNA molecule. Because each of the three bases can be chosen from among the four available—A, C, G, and U—the total possible number of different tRNA molecules should be $4\times4\times4=64$. It is not certain, however, that all 64 are present in the cytoplasm or if they are all necessary.

Each type of tRNA is specific for one or another of the 20 amino acids found in proteins. Because there are 64 possible three-base combinations to specify 20 amino acids, some amino acids are represented by more than one combination. There are, for example, four different combinations, that is, four different tRNA molecules, that specify the amino acid glycine. Some amino acids are represented by one, two, or three different three-base combinations, but a particular combination is specific for one amino acid only.

The function of the tRNA molecules is to pick up from the cytoplasm the amino acids to which they correspond, and to carry them to the ribosome where the mRNA is held. The sequence of nitrogenous bases, taken three at a time (codons), in the mRNA determines the order in which the tRNA molecules, with complementary bases (anticodons), will attach themselves. In this way the mRNA controls both the nature and order of the amino acids in the protein being synthesized (fig. 8.7). After the tRNA molecule has given its associated amino acid to the growing protein chain, it leaves the ribosome and enters the cytoplasm to pick up another amino acid of the same kind, and so on. When the protein is complete, it is free to move away from the mRNA, which can now serve as the tem-

plate for the formation of another protein molecule of the same kind.

The "dictionary" used by the cell to translate from the four-base language of the nucleic acid to the 20-amino-acid language of the proteins is called the genetic code. The translation is performed by the three-base combinations (codons) on the mRNA in conjunction with the corresponding complementary anticodons on the 64 possible tRNA molecules. It is of interest to note that three bases are the minimum number required in a code for translation into the language of 20 amino acids. If the tRNA molecules had only two effective nitrogenous bases, the maximum number of different amino acids that could have been distinguished would have been $4 \times 4 = 16$. Hence, a two-base code would have been inadequate for the purpose of protein synthesis with 20 amino acids. Whether the three-base codon determined the number of amino acids in proteins or the reverse, it is not possible to say.

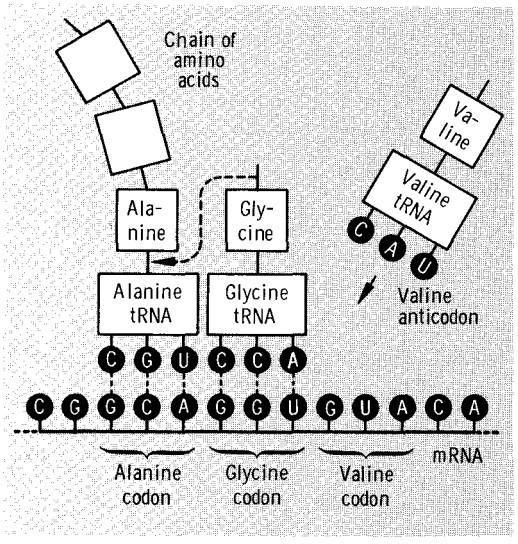


FIGURE 8.7. Formation of a protein by use of messenger RNA and transfer RNA.

Attention may be drawn to a special aspect of the operations in which DNA is involved. Not only does the DNA molecule replicate itself, but it also directs, by means of mRNA, the formation of proteins in the cell. Many of the proteins are enzymes that catalyze the reactions which lead to the production of the individual nucleotides and to their combination to yield nucleic acids. In other words, the DNA controls the formation of the materials (and the energy) required for its own replication.

Nucleic Acids and the Origin of Life

The model described above for the mechanisms of replication and protein synthesis is perhaps oversimplified and undoubtedly incomplete. Nevertheless, there are good reasons for believing that it is substantially correct and that it is applicable to all living organisms. This fact appears to have considerable significance for the origin of life. N. H. Horowitz states:

> The universality of the genetic code, and the related fact that the nucleic acids and proteins of all species are built out of the same nucleotides and amino acids leads to the conclusion that there is really only one form of life on Earth. . . . It is impossible to avoid the conclusion that all species have descended from a common ancestor, which, in the remote past, discovered this remarkably effective and stable genetic mechanism.

As a corollary to the foregoing conclusion, it would appear that the nucleic acids are the key materials for all terrestrial life forms. A nucleic acid, such as DNA, is itself not a living organism, of course, but its presence seems to be fundamental to the essential processes of life. Consequently, it does not seem unreasonable to assume that the origin of life is related to the formation of nucleic acids. This statement should not be taken to imply that nucleic acids alone were involved in the

generation of life. It is probable that polypeptides and perhaps suitable semipermeable membranes were also necessary. Nevertheless, it does not seem that life could have become a continuing and evolving process without substances similar to the nucleic acids.

Some light on the possible mode of production of a primitive DNA may have been shed by the work of the American biochemist Arthur Kornberg and his associates during the latter 1950's. They found that if a mixture containing deoxynucleoside triphosphates,[9] an enzyme (DNA polymerase) extracted from bacterial cells, and a small quantity of a magnesium salt is added to some (primer) DNA outside a living cell, a form of DNA, similar to the primer DNA, is produced quite rapidly. If the primer is absent, however, synthesis of a DNA-like molecule will still occur, although at a very much slower rate. Formation of a type of RNA was observed in similar experiments performed by Severo Ochoa in the United States with nucleoside diphosphates.

It appears certain that enzymes act only as catalysts, accelerating reactions which, in principle, would occur, although perhaps extremely slowly, in their absence. Consequently, as suggested independently by G. W. Beadle and by C. Sagan in 1961, a mixture of appropriate nucleoside phosphates might be expected to produce a polynucleotide, similar to a nucleic acid but possibly simpler, even in the absence of enzymes. An inorganic catalyst, such as the magnesium salt used by Kornberg, might expedite the process.

The rate of formation of nucleic acid might be so slow that it could not be observed in the laboratory. But on the cosmic time scale of hundreds of millions of years, the production of significant amounts of a primitive nucleic acid, even before enzymes existed, might not have been impossible. Once some nucleic acid molecules had been formed, they could presumably serve as primers for the production of more of the same or a complementary type.

A question has been raised concerning the persistence of nucleosides and their phosphates over long periods of time, in spite of the various ways in which such moderately complex molecules might have been decomposed by water, by ultraviolet radiation, and so on. A possible answer may be found in a consideration of some aspects of the behavior of enzymes. There is experimental evidence that the whole of an enzyme molecule is not required to catalyze a particular reaction. For example, it has been shown in some cases that much of the catalytic activity occurs at a site consisting of a relatively small number of amino acids. Furthermore, two-thirds of the approximately 180 simple amino acids in the protein-digesting enzyme papain can be removed without affecting its catalytic activity.

It is conceivable that a relatively simple peptide, containing only a few amino acids, could have been present on Earth even before life developed. Such a peptide might have served as a catalyst, even if not a very good one, for the linking of nucleoside phosphate molecules to form a polynucleotide chain. There is actually laboratory evidence that synthetic peptides do, in fact, possess weak catalytic activity for certain reactions, although not yet for the polymerization of nucleotides.

Incidentally, the formation of a peptide, or polypeptide, and a polynucleotide need not have been independent. A popular explanation of the specificity of each enzyme for a

[9] A deoxynucleoside triphosphate is a combination of a nitrogenous (pyrimidine or purine) base, deoxyribose, and three phosphate residues. The molecule thus has two phosphate residues more than a deoxynucleotide.

particular reaction is that the shape of the protein (enzyme) molecule is such as to permit the attachment of certain suitably shaped molecules, or parts of molecules, only. It is quite possible, or even probable, therefore, that the latter influenced in some manner the production of the enzyme catalyst. Thus, a polynucleotide chain may have favored the development of a polypeptide catalyst capable of facilitating the formation of similar chains. In the course of time, better and better catalysts would have arisen, and because they used the available material more effectively, they would tend to predominate over the others by a process akin to evolution by natural selection among living organisms.

Another circumstance which may have played a role in the synthesis of nucleic acids and proteins on the primitive Earth is the catalytic effect of solid surfaces. The action of suitable solids in catalyzing many reactions, both organic and inorganic, in the laboratory and in industry is, of course, well known. Adsorption on mineral and clay surfaces might have favored the linking up of nucleotides into chains in preference to reactions which would have led to their destruction. Polypeptide formation from available amino acids could have occurred in an analogous manner. Ions in solution are also capable of acting as catalysts. Thus, in the production of DNA from a mixture of deoxynucleoside triphosphates, mentioned above, the presence of magnesium ions seems to be required.

Some General Considerations

Of course, there is an enormous gap between nucleic acids and polypeptides, on the one hand, and their organization into a living system capable of replication, mutation, and production of the specific proteins necessary for its operation, on the other hand. "A major unsolved problem" writes C. Sagan, "is the origin of the genetic code. The code operates today only in the presence of . . . ribosomes, activating enzymes, transfer RNA, etc., which could not be expected at the time of the origin of life. How in earlier times did a sequence of nucleotides in a nucleic acid determine a sequence of peptides in a polypeptide?"

A review of the progress of natural and biological sciences during the past 100 years would suggest that this and other problems will be solved and that the gap will eventually be bridged. It will be assumed for the present, therefore, that if polynucleotides and polypeptides had been available, in suitable circumstances, at some early stage of Earth's history, life, as it now exists, could have developed as the result of many stages of evolution. The fact that nothing definite is known about the intermediate steps should not preclude the reasonable assumption that they did occur in the distant past.

The suggestion has been made that viruses, which are larger than a single molecule but smaller than a cell, or its nucleus, represent an intermediate stage between nonliving and living systems. A virus consists mainly of a nucleic acid, DNA or RNA, surrounded by protein. It cannot replicate on its own, but will do so when introduced into a living cell. The virus then uses the machinery of the cell—the nucleotides, amino acids, and enzymes—to reproduce its characteristic nucleic acid and proteins. Because viruses are unable to replicate without living organisms, it would appear, at first sight, that they could not have evolved before living cells. It is conceivable, however, that primitive viruses originally had a crude means of reproduction, but with the evolution of living cells, they became adapted to a superior mode of replication and the old one died out.

319-889 O—69——6

Some biologists feel that, although nucleic acids and proteins are intimately related to the origin of life, another factor was essential. In 1963, R. S. Young said:

> Most of the complex compounds of the cell must be present in the medium in adequate concentrations [for replication to occur] . . . and it hardly seems likely that this would have been the composition of the early ocean. . . . It seems much more likely for the first replicating system to have had a specialized environment in which it was able to assemble the required materials and keep out those which were deleterious. . . . This is accomplished by cells today by means of a cell membrane, which is a selectively permeable barrier composed of protein and lipid. It permits the cell to create its own internal environment necessary for its metabolic and reproductive activity. It is difficult to imagine a living unit, no matter how primitive or simple, without some such capability.

In other words, it is considered that a suitable membrane, to form the equivalent of a cell, was necessary before life could have originated.

The earliest known fossils of living organisms, which are moderately complex, are estimated to be more than 3 billion years old. It may be guessed, therefore, that life has existed on Earth for some 3.5 to 4 billion years. The age of Earth, that is, the time since the body of Earth developed into approximately the same form as at present, is considered to be 4.6 billion years. Hence, approximately from half a billion to a billion years elapsed before the most primitive forms of life, perhaps not recognizable as life today, evolved on Earth from the material available at the time.

Attempts have been made, notably by A. I. Oparin, to indicate the possible stages of evolution from nonliving to living matter, but these are no more than speculation based on the limited available knowledge. Nevertheless, it would not seem unreasonable to assume that if there had been present on the primitive Earth, such substances as pyrimidines, purines, and five-carbon sugars for the formation of polynucleotides, and amino acids for the production of proteins, as well as a means for providing a suitable environment, life would have developed in the course of time.

FOR FURTHER INFORMATION

ASIMOV, I.: The Genetic Code: Signet Science Library, The New American Library, Inc., 1962.

BERNAL, J. D.: The Origin of Life. World, 1967.

CRICK, F. H. C.: The Structure of the Hereditary Material. Sci. Am., vol. 191, no. 4, 1954, p. 54.

CRICK, F. H. C.: The Genetic Code. Sci. Am., vol. 207, no. 4, 1962, p. 66.

CRICK, F. H. C.: The Genetic Code: III. Sci. Am., vol. 215, no. 6, 1966, p. 55.

FOX, S. W.: The Origin of Life. Nature, vol. 205, 1965, p. 328.

HOROWITZ, N. H.: The Biological Significance of the Search for Extraterrestrial Life. Advan. Astronaut. Sci., vol. 22, 1967, p. 3.

KEOSIAN, J.: The Origin of Life. Reinhold, 1968.

NIERENBERG, M. W.: The Genetic Code: II. Sci. Am., vol. 208, no. 3, 1963, p. 80.

OPARIN, A. I.: Life: Its Nature, Origin, and Development. English translation, Academic Press, 1961.

OPARIN, A. I.: The Origin and Initial Development of Life. NASA TT F–488, 1968.

SAGAN, C.: Radiation and the Origin of the Gene. Evolution, vol. 11, 1957, p. 40.

SAGAN, C.: On the Origin and Planetary Distribution of Life. Radiation Res., vol. 15, 1951, p. 174.

SHKLOVSKII, I. S.; AND SAGAN, C.: Intelligent Life in the Universe. Holden-Day, 1966.

SHNEOUR, E. A.; AND OTTESEN, E. A., COMPS.: Extraterrestrial Life: An Anthology and Bibliography. Natl. Acad. Sci.-Natl. Res. Council Publ. 1296A, 1966.

SULLIVAN, W.: We Are Not Alone. McGraw-Hill, 1964.

WALD, G.: The Origin of Life. Sci. Am., vol. 191, no. 2, 1954, p. 44.

IX

The Formation of Prebiological Materials

PRODUCTION OF POSSIBLE PREBIOLOGICAL MATERIALS

Earth's Primitive Atmosphere

PREBIOLOGICAL MATERIALS are defined as various organic (carbon) compounds from which complex molecules, such as nucleic acids and proteins, present in living organisms evolved during the first billion years or so of Earth's history. As seen in the preceding chapter, the most important prebiological compounds are probably alpha amino acids, certain pyrimidines and purines, and the five-carbon sugars, ribose and deoxyribose. Except for the few amino acids which contain sulfur atoms, the only elements present in the compounds mentioned are hydrogen, carbon, nitrogen, and oxygen. The nucleotides, of course, also contain phosphorus.

The first attempts to simulate the production of prebiological compounds in the laboratory were largely failures. A reasonable explanation of this lack of success was proposed by H. C. Urey in 1952; namely, that the assumptions made concerning Earth's primitive atmosphere were probably incorrect. Subsequently, many experiments have been made with mixtures of gases that are considered to be representative of the composition of Earth's atmosphere some 4 billion or more years ago. As a result, considerable progress has been made in the production of substances that fulfill the requirements of a prebiotic system.

For many years it has been taken for granted that Earth's atmosphere had always contained some free oxygen gas, although probably not as much as it does now (21 percent by volume). Carbon compounds in the atmosphere would then be present mainly as carbon dioxide. In the course of his detailed development of a mechanistic theory of the origin of life, first stated in 1924 and elaborated in 1936, as mentioned in chapter VIII, A. I. Oparin cast doubt on the previously accepted composition of Earth's early atmosphere.

Largely on the basis of geochemical evidence, Oparin suggested that the carbon in the atmosphere existed originally in the reduced form as hydrocarbons and not in the oxidized form as carbon dioxide. Similarly, nitrogen, now present in the free state in the air, would have been in the reduced form of

ammonia. Oparin thought that in this atmosphere, containing hydrocarbons such as methane (CH_4), in addition to ammonia (NH_3) and water (H_2O) vapor, amino acids and other prebiological compounds might have been formed in the early stages of Earth's history. Subsequent experimental studies, to be described later, have shown that several amino acids, among other substances, can indeed be produced by supplying energy in a convenient form to a reducing atmosphere of the type envisaged by Oparin.

Independently, J. B. S. Haldane had been considering the question of the composition of Earth's atmosphere during the period that preceded the development of life. Like Oparin, but on entirely different grounds, Haldane concluded in 1928 that the atmosphere could have contained little or no free oxygen. His arguments were based on biological considerations relating to the metabolic processes occurring in anaerobic and aerobic organisms. The anaerobic systems are able to live and multiply in the complete absence of oxygen gas; in fact, some bacteria, called obligate anaerobes, are unable to grow if any free oxygen is present. For aerobic systems, on the other hand, oxygen is essential.

One of the points made by Haldane was that the anaerobic stages of intermediate metabolism are common to all organisms on Earth, whereas the aerobic steps show considerable diversity from one organism to another. He concluded, therefore, that anaerobic forms of life had evolved before the aerobic ones and that life had originated in an atmosphere that contained essentially no free oxygen gas. Haldane thought that Earth's primitive atmosphere consisted mainly of carbon dioxide, ammonia, and water vapor. The carbon was thus considered to be in the form of the dioxide rather than as a hydrocarbon, as Oparin had suggested.

Relatively little of the ultraviolet light from the Sun is able to penetrate Earth's present atmosphere because most of it is absorbed by oxygen (O_2) and ozone (O_3), the latter being formed from oxygen. Absorption of the radiation results in decomposition of these substances, but the oxygen atoms produced can recombine to regenerate oxygen and ozone molecules. In an atmosphere that contained no oxygen, the other constituents would absorb the ultraviolet light and undergo chemical reaction. Haldane suggested that such radiation acting on a mixture of carbon dioxide, ammonia, and water vapor in Earth's primitive atmosphere would have produced "a vast variety of organic substances . . . including sugars and apparently some of the materials from which proteins are built up." These substances would then have accumulated in the oceans to form the "hot dilute soup" from which life could have originated.

The speculations of Oparin and of Haldane did not arouse much interest among experimentalists, but the situation underwent a change after 1952. In considering the problem of the origin of the planets, mainly from a physicochemical point of view, H. C. Urey showed on thermodynamic grounds that Earth's primitive atmosphere would be expected to consist of methane, ammonia, water vapor, and some free hydrogen gas.

Urey's conclusions have been disputed by the American geophysicist P. H. Abelson. In an article published in 1966, the latter contends that the atmosphere originally contained mainly carbon monoxide, carbon dioxide, nitrogen, and hydrogen. Ultraviolet radiation (or an electrical discharge) acting on such a mixture of gases would produce appreciable quantities of hydrogen cyanide (HCN). The latter would then react in various ways to form compounds of biological significance. The system would not be in thermodynamic equilibrium, but this is conceivable because of the

slow rate of attainment of equilibrium and the continuous release of gases, such as nitrogen and oxides of carbon, from Earth's interior.

For the immediate problem, however, the exact composition of Earth's early atmosphere is not critical, especially as hydrogen cyanide can be formed from a variety of gas mixtures. The main requirement is that it should be a reducing atmosphere. That is to say, methane, carbon monoxide, ammonia, and hydrogen must predominate over carbon dioxide and nitrogen gas. Fortunately, there is fairly general agreement on this point. It appears that differences in the actual composition of the atmosphere would not affect significantly the nature of the prebiotic compounds formed, although the chemical mechanism of their formation might be different.

Before leaving the subject of Earth's atmosphere, some further remarks will be made about atmospheric oxygen. The consensus is that the primitive reducing atmosphere changed gradually to one with oxidizing character. This may have resulted from the decomposition of water vapor into hydrogen and oxygen gases by ultraviolet light. The hydrogen, being a very light gas, would escape from Earth's gravitational field, whereas most of the oxygen would be retained. The reduced compounds in the atmosphere would first be oxidized and then free oxygen gas would be available. Some of this gas appears to have been used up in the oxidation of minerals, but about a billion years ago free oxygen began to accumulate in the atmosphere.

With the increasing atmospheric content of oxygen and ozone, there was a marked decrease in the amount of solar ultraviolet radiation reaching Earth's surface. As a result, living organisms were able to leave the oceans, where the water had protected them from the harmful effects of this radiation, and settle on land. Green plants, which were able to utilize the carbon dioxide in the atmosphere and produce oxygen gas by the process of photosynthesis in sunlight, then developed rapidly. Consequently, the amount of oxygen in Earth's atmosphere began to increase at a considerable rate, reaching the relatively high proportion of the present day.

Formation of Amino Acids

One of the consequences of Urey's ideas concerning the composition of Earth's primitive atmosphere was that his associate S. L. Miller, at the University of Chicago, started a series of laboratory studies on the formation of prebiological substances. These studies not only led to important results, first published in 1953, but they stimulated others to undertake experiments that might throw light on the origin of compounds which are considered to have been essential in the evolution of living organisms from inorganic materials.

There were several sources of energy available for causing chemical reactions to occur among the gases in the primitive atmosphere. These included radiations, especially ultraviolet, from the Sun; electrical discharges (lightning); cosmic rays (from the Sun and space); radioactivity from minerals in Earth's outer layer; and local high temperatures (volcanoes). An analysis made by H. C. Urey and S. L. Miller indicates that solar ultraviolet light was probably the most important source of energy, with electrical discharges next in order. In simulating primitive Earth conditions, it would have been desirable to use ultraviolet radiation as the energy source, but because of the experimental difficulties, an electrical discharge was employed by Miller in his pioneering work. Later investigations made by others have shown that the results obtained in this manner are quite similar to those from experiments using ultraviolet light.

The experimental arrangement employed by Miller, to simulate Earth's early atmos-

phere and bodies of water, is shown in figure 9.1. A mixture of hydrogen, methane, and ammonia gases, in the proportions of 1 to 2 to 2 by volume, at a total pressure of 50 centimeters of mercury, was introduced into glass bulb *A*. By means of two electrodes inserted through the wall, an electrical discharge, at a peak potential of 60 000 volts, could be passed through the gases by means of a Tesla coil. Flask *B* contained water which was heated to boiling, so that water vapor passed up the tube at the left and entered bulb *A*. A small quantity of water in U-tube *C* prevented circulation in the opposite direction. The products of the spark discharge were condensed at *D* and flowed through the U-tube back to flask *B* where they accumulated.

After the experiment had been run for about a week, the liquid in the flask was removed and subjected to analysis. The products included several amino acids as well as other organic compounds of biological interest. The amino acids will be discussed here, whereas the other substance will be considered in later sections. The amino acid obtained in largest yield was glycine (p. 162), but there were also formed significant quantities of alanine, aspartic acid, and glutamic acid, which are present in proteins. In addition, a few amino acids that do not occur in proteins but are closely related to them were produced. The same substances, although in somewhat different actual and relative amounts, were formed by passing a silent electrical discharge from an ozonizer through the mixture of gases.

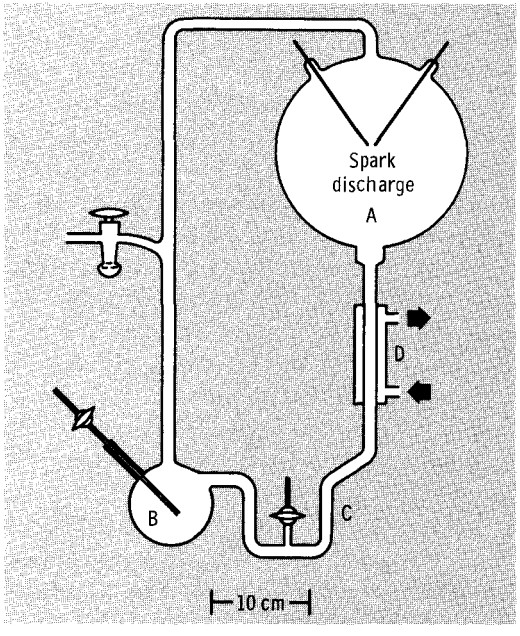


FIGURE 9.1. Apparatus for simulating an electrical discharge in Earth's primitive atmosphere. (After S. L. Miller.)

To follow the progress of the reactions in the electrical discharge, small samples were taken from the U-tube and from flask *B* at regular intervals and analyzed for the ammonia remaining and for hydrogen cyanide, aldehydes, and total amino acids formed. The results are given in figure 9.2. The continuous decrease in the quantity of ammonia is mainly caused by decomposition rather than by the formation of amino acids. Incidentally, the presence of hydrogen cyanide (HCN) in the solution containing ammonia (NH_3) indicates that part of the cyanide exists as ammonium cyanide (NH_4CN).

Passage of an electrical discharge or ultraviolet radiation of short wavelength through the mixture of methane, ammonia, water vapor, and hydrogen undoubtedly leads to the formation of various free radicals and hydrogen atoms. Examples are CH_3, CH_2, and CH from methane, NH_2 and NH from ammonia, OH from water, and H atoms. Interaction among these radicals or atoms would be expected to lead to the formation of hydrogen cyanide and aldehydes, as well as other substances.

The presence of hydrogen cyanide and aldehyde among the early products of the

electrical discharge reactions suggests that amino nitriles, such as are known to be formed by the interaction of ammonia, hydrogen cyanide, and aldehyde (RCHO), are intermediates in the production of amino acids. The first step may be represented by

$$NH_3+HCN+RCHO\rightarrow NH_2CH(R)CN+H_2O$$

Amino nitrile

and the amino nitrile then undergoes reaction with water, in two stages. The first stage leads to the formation of an amino acid amide by the reaction

$$NH_2CH(R)CN+H_2O\rightarrow NH_2CH(R)CONH_2$$

Amino acid amide

and the second stage yields an alpha amino acid by hydrolysis:

$$NH_2CH(R)CONH_2+H_2O\rightarrow NH_2CH(R)COOH+NH_3$$

Alpha amino acid

This process for the formation of alpha amino acids is known as the Strecker synthesis. It has been suggested that the production of amino acids by an electrical discharge in the mixture of methane, ammonia, and water vapor takes place, to a large extent, by this mechanism.

Formaldehyde (HCHO) is known to be formed in the electrical discharge. In this case R in the general formula given above for aldehyde is the hydrogen atom (H). The corresponding amino nitrile is NH_2CH_2CN (amino acetonitrile), the amino acid amide is $NH_2CH_2CONH_2$ (amino acetamide), and the amino acid is NH_2CH_2COOH (amino acetic acid or glycine). Similarly, when the aldehyde is acetaldehyde (CH_3CHO), R is CH_3, and the amino acid produced is $NH_2CH(CH_3)COOH$, which is alpha alanine.

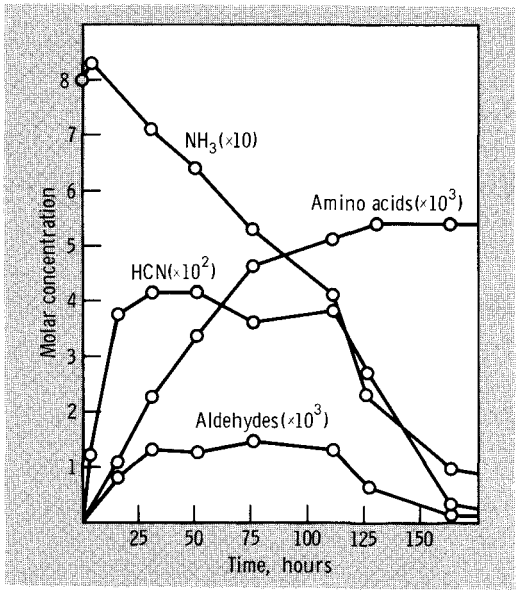


FIGURE 9.2. Products of electrical discharge through mixture of gases. (After S. L. Miller.)

From the experiments of P. H. Abelson (1956) in the United States and of T. E. Pavlovskaya and A. G. Pasinsky (1957) in the U.S.S.R., it has been established that amino acids can be produced by the passage of an electrical discharge through various gas mixtures other than those used by Miller in the work just described. Abelson, for example, studied various mixtures of gases selected from hydrogen, methane, carbon monoxide and dioxide, ammonia, nitrogen, water vapor, and oxygen, whereas the Russian scientists used a mixture of methane, carbon monoxide, ammonia, and water vapor with no free hydrogen.

Provided the gaseous system has overall reducing properties, amino acids are always formed. If oxygen is added in an amount sufficient to offset the reducing character of such gases as hydrogen, methane, carbon monoxide, and ammonia, amino acids are not

obtained. It is of interest to mention in this connection that, as early as 1913, W. Loeb in Germany had observed the formation of amino acids when a silent electrical discharge was passed through a mixture of water vapor with carbon monoxide and ammonia, obtained by decomposition of formamide ($HCONH_2 \rightarrow CO + NH_3$).

The production of amino acids from mixtures of appropriate gases with reducing properties, with ultraviolet light as the energy source, was reported in 1957 by W. Groth and H. von Weyssenhoff in Germany, by A. N. Terenin in the U.S.S.R., and later by others. The gas mixtures were methane, ammonia, and water vapor with or without the addition of carbon monoxide. The products were quite similar to those obtained with the silent electrical discharge, but the yields were lower. A possible explanation is that with ultraviolet radiation less hydrogen cyanide is formed, as an intermediate stage, than with an electrical discharge. N. H. Horowitz and S. L. Miller have suggested that, in prebiological times, "the electrical discharges were [possibly] the principal source of hydrogen cyanide and that ultraviolet light was the principal source of aldehydes, so that the two processes complemented each other." The important point is that similar final products are obtained regardless of whether the energy is supplied by an electrical discharge or by ultraviolet radiation.

Another possible source of energy is accelerated electrons, which are equivalent to the beta particles from radioactive materials, such as potassium-40 present in many minerals (p. 75). Studies performed by C. Palm and M. Calvin at the University of California in 1962 and by Juan Oró at the University of Houston in 1963 have shown that amino acids also result from the irradiation of mixtures of methane, ammonia, and water vapor by high-energy electrons. It would appear, therefore, that the nature of the energy source is not important. Amino acids would have formed in a reducing atmosphere, containing gaseous compounds of carbon, nitrogen, oxygen, and hydrogen, regardless of the particular source, or sources, of energy available.

Reactions in Solution

In the experiments described above, the starting materials were gases in every case. It is probable, however, that the later stages of the reactions leading to the formation of amino acids actually took place in aqueous solution. Consequently, studies of solutions containing materials which could have been formed in the primitive atmosphere are pertinent to the present issue. The Russian scientists Pavlovskaya and Pasinsky exposed to ultraviolet light an aqueous solution containing 2.5 percent of formaldehyde together with 1 to 1.5 percent of ammonium chloride or nitrate. Among the products were the amino acids glycine, alanine, penylalanine, glutamic acid, valine, serine, and isoleucine, which are found in proteins. The concentration of formaldehyde used in these experiments was considerably higher than would be expected on the primitive Earth. It is probable, however, that the same reactions would occur, although much more slowly, in less-concentrated solutions of formaldehyde.

From the results of the analyses performed by Miller during the course of his work (fig. 9.2), there is no doubt that ammonium cyanide was present in the condensate solution. Several amino acids have been obtained by J. Oró (1961) and by C. V. Lowe (1963) and their respective collaborators by heating aqueous solutions of ammonium cyanide for long periods of time. Again, however, the concentrations of ammonium cyanide were much greater than would seem reasonable under primitive Earth conditions. It should be noted that in these experiments starting with

ammonium cyanide, heat was the only source of energy used.

If, as suggested earlier, free radicals are produced in the primary stage of the action of ultraviolet radiation or an electrical discharge on the methane, ammonia, and water vapor (or equivalent) mixture, then, according to J. Oró, such substances as a hydroxylamine (NH_2OH from NH_2 and OH radicals) and hydrazine (N_2H_4 from two NH_2 radicals) should be formed. Consequently, Oró and his coworkers (1959) studied the reaction between formaldehyde and hydroxylamine in various solutions at temperatures of 70° to 100° C and identified several amino acids among the products. Formation of these substances could thus have occurred on the primitive Earth by this mechanism, as well as by the Strecker synthesis involving amino nitriles.

High-Temperature Reactions

Another approach to the problem under consideration has been taken by the American chemist Sidney W. Fox and his associates since 1958. Fox has contended that reactions involving only gases of the atmosphere, without liquid water, were also important in the synthesis of amino acids. The energy for these reactions would have been supplied as heat from the high-temperature (volcanic) regions that were presumed to have existed in many parts of Earth's crust.

In support of this view, Fox quotes the results of his experiments in which a gaseous mixture of methane, ammonia, and water vapor was passed through a tube containing silica sand, silica gel, or volcanic lava heated to a temperature of 900° to 1100° C. The vapors formed were condensed in water, and analysis showed the presence of nearly all the amino acids occurring in proteins, with the exception of those having sulfur atoms in the molecule. Presumably high-temperature reactions of this kind could have taken place at local hot regions in the primitive Earth's crust, with the amino acids condensing in an adjacent cooler area.

Formation of Polypeptides

There are evidently several fairly plausible ways in which amino acids could have formed from a reducing atmosphere on Earth as one of the preliminary stages that led to the development of life. It will now be shown that, once the amino acids were available, polymerization into peptide (and polypeptide) chains, the precursors of proteins, could have occurred under suitable conditions. Several methods for producing peptides from simple amino acids have been developed in the laboratory, but the only procedures to be described here are those which might conceivably have occurred on the primitive Earth.

In the work referred to earlier, S. L. Miller tested for peptides among the products he obtained by the passage of an electrical discharge through the methane, ammonia, water vapor mixture, but the results were not decisive. Later, others repeated the experiments and did detect peptides, but in small amounts. Thermodynamic data indicate that, in aqueous solution at normal temperatures, the polymers should be unstable relative to the simple amino acids. Consequently, large quantities of peptides would not be expected under the conditions of the electrical discharge experiments.

Theoretical considerations, however, suggested that at higher temperatures, increased peptide formation should occur. This was confirmed by J. Oró in 1961 by heating glycine in an aqueous solution of ammonia at temperatures up to 140° C under pressure. Polypeptides containing as many as 18 glycine residues were isolated from the products. If other amino acids were added to the glycine

solution in aqueous ammonia, polypeptides containing at least eight different amino acids were obtained upon heating.

If amino acids are formed in the simulated primitive Earth atmosphere by the Strecker synthesis, then amino nitriles and amino amides would be intermediate products. Solutions in water of substances of these two types have been converted into peptides by heating, especially in the presence of ammonia. Thus, peptides were undoubtedly formed in Miller's experiments, and their presence in similar experiments was subsequently established by others.

Amino acids have been obtained by heating aqueous solutions of ammonium cyanide. Polymers containing several different amino-acid residues are formed at the same time, but it does not seem to have been established that these are true polypeptides; that is, molecules with a —N—C—C—N—C—C— chain.

According to the studies made by S. Akabori and his colleagues in Japan since 1956, aminoacetonitrile can be converted into simple polymers of glycine by heating in water in the presence of clay which serves as a catalyst. If formaldehyde or acetaldehyde is added subsequently and the medium is made alkaline, some of the glycine residues are converted into other amino acids, such as serine and threonine, respectively, with the clay again acting as a catalyst. Formaldehyde thus causes the substitution of the side chain $—CH_2OH$ and acetaldehyde the side chain $—CH(CH_3)OH$ for a hydrogen atom on the alpha carbon atom of glycine.

The formation of simple peptides has been demonstrated in dilute solution with the necessary energy supplied by radiation. In 1965, the Ceylon-born Cyril Ponnamperuma and his associate Etta Peterson, working at the Ames Research Center of the National Aeronautics and Space Administration, exposed a dilute solution of the amino acids glycine and leucine to ultraviolet light in the presence of cyanamide as a catalyst. Cyanamide, which has the formula $CN \cdot NH_2$, is undoubtedly formed in the early stages of the simulated primitive atmosphere reactions. Among the products detected were four dipeptides and one tripeptide. Similarly, M. Calvin and his associates found that a dipeptide was formed in a dilute solution of alanine if dicyanodiamide, a dimer of cyanamide, was added, even without ultraviolet irradiation.

Polypeptide Formation in Nonaqueous Systems

Because the formation of a polypeptide from simple amino-acid molecules involves the elimination of water, it is to be expected that polymerization would take place more readily in the absence of water. This line of investigation has been followed by S. W. Fox and his collaborators, who have studied the thermal polymerization of amino acids by heating dry samples at temperatures of 150° to 200° C for several hours. When leucine, phenylalanine, or tryptophan was heated alone, the product was not a polypeptide. Mixtures of aspartic acid and leucine, of glutamic acid and phenylalanine, and of lysine and alanine did yield polypeptides, however.

Especially successful was an experiment in which 16 of the common amino acids were heated with excess of both aspartic acid and glutamic acid. The two latter are unusual in the respect that they are dicarboxylic acids containing two —COOH groups. The polymers obtained in this manner were found to show a remarkable resemblance to proteins in many respects, as will be seen below. They have, consequently, been called proteinoids. "The term proteinoid," writes S. W. Fox, "is

employed to indicate a polymeric preparation containing, in peptide linkage, all or most of the amino acids common to protein."

The mean molecular weights of proteinoids obtained by heating amino-acid mixtures at 160° C was found to be about 3600, whereas heating at 190° C gave an average around 8600. These values may be compared with the molecular weight of insulin, which is approximately 6000. It should be noted that the molecular weights of the proteinoids quoted are mean values of a complex mixture of polypeptides. Some components may have considerably larger molecular weights.

Proteolytic enzymes, such as trypsin, chymotrypsin, pepsin, and papain, which catalyze the splitting of protein molecules in aqueous solution, are able to decompose proteinoids in a similar manner. This is one example of the resemblance between proteinoids and proteins. Furthermore, the proteinoids are found to have nutritive value for bacteria when substituted for peptone, which is a partially digested protein.

There are other similarities to natural proteins, but there is at least one apparent difference. The synthetic proteinoids do not produce the antigen reaction in rabbits and guinea pigs such as would result from introduction of a foreign protein. A possible reason is that the molecular weight of the proteinoid is not high enough. As a general rule, only proteins with molecular weights of at least 10 000, and possibly 15 000, are capable of causing the antigen reaction. A striking property of proteinoids, which may have some relevance in connection with the origin of life, will be considered later in this chapter.

S. W. Fox has stated, "Temperatures above 150° C cannot be ruled out as one of the primordial conditions." Some scientists agree with this view, but others have considered it improbable that the circumstances could ever have permitted dry amino acids to be heated to such temperatures for a sufficient length of time to produce polypeptides of high molecular weight. The results of the thermal polymerization experiments are nevertheless of interest. Polypeptides analogous to those obtained by heating at 150° C for a few hours in the laboratory might conceivably have been formed on the primitive Earth at lower temperatures over a longer period of time, especially in the presence of a catalyst.

Although their nature is unknown, catalysts could have played an important part in the polymerization of amino acids to form peptides, polypeptides, and proteins. For example, since 1958, Gerhard Schramm and his colleagues in Germany have used a material called metaphosphate ester or ethyl metaphosphate as a condensing agent, which is effectively a catalyst for the elimination of molecules of water (p. 163), to produce polypeptides at moderate temperatures. The metaphosphate ester is made by heating phosphorus pentoxide with diethyl ether in chloroform solution. In the presence of this substance, the tripeptide alanylglycylglycine was converted into a mixture of polymers containing from 24 to 40 amino-acid residues. By a similar procedure, a polypeptide with a molecular weight around 7300 was obtained from a mixture of the three simple amino acids: tyrosine, alanine, and glutaminic acid.

Other catalysts (or condensing agents) derived from phosphoric acid were used in 1963 by K. Harada and S. W. Fox to polymerize amino acids. One such substance was a polyphosphoric acid made by heating common orthophosphoric acid at temperatures from 200° to 350° C. When aspartic acid, alone or mixed with another amino acid, was heated with the polyphosphoric acid at temperatures as low as 70° C for a few days, polypeptides with molecular weights as high

as 13 000 to 15 000 were produced. A more complex proteinoid was obtained by heating a mixture of glutamic and aspartic acids and 14 other amino acids with polyphosphoric acid at 100° C for 4 to 6 days. By using ethyl metaphosphate as the condensing agent, R. S. Young was able to obtain proteinoids at temperatures as low as 25° C.

It would not seem to be very probable that such concentrated substances as ethyl metaphosphate or polyphosphoric acid were involved in the production of prebiological compounds on a primitive Earth. The results are nevertheless of interest as indicating what is possible when suitable condensing agents are used to facilitate the polymerization of amino acids. It may perhaps be significant that phosphates are involved in many important biological compounds, including the nucleic acids and adenosine triphosphate, and in biological processes, such as photosynthesis.

Formation of Pyrimidines and Purines

Because of the undoubted significance of nucleic acids in life processes, consideration will now be given to the possible prebiological synthesis of pyrimidines and purines, the nitrogenous bases of the nucleic acids. Evidence has been obtained that the two purines, adenine and guanine, and two of the three pyrimidines in nucleic acids, cytosine and uracil, could have been formed under primitive Earth conditions.

Among the products obtained by passing an electrical discharge through a reducing atmosphere is the substance urea, $CO(NH_2)_2$. It is also formed, as will be seen shortly, by the action of ultraviolet radiation on a dilute solution of hydrogen cyanide. Furthermore, one of the amino acids produced in substantial amounts is beta-alanine ($NH_2CH_2CH_2COOH$), which does not generally occur in natural proteins. Intermediate stages in the formation of this acid are probably acrylonitrile ($CH_2{:}CHCN$), beta-amino propionitrile ($NH_2CH_2CH_2CN$), and beta-amino propionamide ($NH_2CH_2CH_2CONH_2$). J. Oró has found that if any of these latter three substances is heated together with urea in aqueous solution containing ammonium chloride at a temperature of 135° C, a small quantity of uracil will be found among the products.

Chemists have known for many years that uracil can be synthesized in the laboratory by heating a mixture of malic acid, urea, and sulfuric acid. In 1961, S. W. Fox and K. Harada found that by substituting polyphosphoric acid for the sulfuric acid and heating at 100° to 240° C for about an hour, uracil could be obtained. Aspartic acid is one of the substances produced when an electrical discharge or ultraviolet radiation is passed through a simulated primordial atmosphere, and malic acid could be formed from it by the action of water. The two acids differ only in the respect that the amino (NH_2) group in the former is replaced by a hydroxyl (OH) group in the latter.

As noted above, however, it is improbable that the highly concentrated polyphosphoric acid would exist on the primitive Earth, but there might be other substances which have the same effect but to a lesser extent. It is of interest, therefore, that small amounts of uracil were obtained in a less drastic manner by J. Oró in 1963 by heating an aqueous solution of acrylonitrile, urea, and ammonium chloride to 135° C.

A more plausible approach to the prebiotic synthesis of the pyrimidines cytosine and uracil is indicated by the work of L. E. Orgel and his associates at the Salk Institute of Biological Studies in 1968. They found that good yields of cytosine resulted from the heating at 100° C of a relatively dilute aqueous solution of cyano-

acetylene ($HC \equiv C \cdot CN$) and ammonium cyanate (NH_4CNO). Both of these substances could have been formed in a reducing atmosphere containing nitrogen, carbon, and hydrogen compounds in the presence of water. Cytosine is readily converted into uracil by hydrolysis, in which a molecule of ammonia is removed and a molecule of water is added. The third pyrimidine found in nucleic acids, thymine, has not yet been produced under simulated primitive Earth conditions.

The purine adenine, which is in a sense a fivefold polymer of hydrogen cyanide, $(HCN)_5$, was found by J. Oró and A. P. Kimball in 1961 to be present in appreciable quantities in a solution of hydrogen cyanide and ammonia after heating for a few days at 70° C. In addition, several precursors (or intermediate compounds) were isolated. These are of special interest because of their role in the formation of other substances of biological significance.

Among the intermediates in the formation of adenine was the compound known as 4-amino imidazole 5-carboxamide. Upon heating this in aqueous solution with either urea or guanidine, the second purine present in nucleic acids, guanine, was obtained. Guanidine is among the substances formed when an aqueous solution of ammonium cyanide is heated. It is related structurally to urea, the O atom in the latter being replaced by the equivalent NH group to give $HN=C(NH_2)_2$. The function of urea or guanidine is evidently to add a carbon atom to the carboxamide compound and thereby permit the formation of the six-membered (pyrimidine) ring in guanine.

In 1963, C. Ponnamperuma and Ruth Mariner reported the formation of both adenine and guanine under conditions much closer to those expected for the primitive Earth. They did this by exposing a very dilute solution of hydrogen cyanide to ultraviolet light for several days. It is noteworthy that substantial quantities of urea, which may have been involved in the formation of uracil, were produced at the same time.

C. Ponnamperuma and his associates were also able to synthesize adenine, but apparently not guanine, by exposing a mixture of methane, ammonia, and water vapor, with or without hydrogen gas, to high-energy (4.5 million electron volts) electrons. These electrons were intended to simulate the beta particles emitted in the radioactive decay of potassium-40. The yields of adenine were larger in the absence of free hydrogen than in its presence. This is not altogether unexpected because the carbon atoms in adenine are in a higher state of oxidation than in methane. The failure to detect guanine may be an indication of the greater stability of adenine. In addition to its presence in nucleic acids, the latter occurs in other compounds of biological importance, such as adenosine mono-, di-, and tri-phosphates.

It is perhaps somewhat surprising that guanine was formed, although in only a small amount, in a more drastic process. This substance was found in 1964 by C. Ponnamperuma and his coworkers in the product obtained by heating at 180° to 200° C the mixture of 18 amino acids used by S. W. Fox to prepare proteinoids. None of the other four nitrogenous bases was detected.

A result of exceptional interest in connection with the prebiotic formation of pyrimidines and purines was reported in 1968 by M. H. Studier, E. Anders, and R. Hyatsu of the University of Chicago and the Argonne National Laboratory. They heated a mixture of carbon monoxide, hydrogen, and ammonia gases in the presence of a catalyst containing iron, nickel, and cobalt obtained from a meteorite (p. 196). Adenine, guanine, and

cytosine were detected among the products, and the view was expressed that thymine would probably be found in due course.

Formation of Sugars

In addition to pyrimidines and purines, nucleic acids contain the five-carbon sugars ribose (in RNA) and deoxyribose (in DNA). These substances, which are compounds of hydrogen, carbon, and oxygen, but not nitrogen, should have been available under prebiological conditions. It has been known since 1861 that various sugars, all of which are represented by the general formula $(CH_2O)_n$, where n is 5 for ribose, are formed in an alkaline solution of formaldehyde (CH_2O). The base used to make the medium alkaline evidently acts as a catalyst for the condensation (or polymer formation) of three or more molecules of formaldehyde.

As already seen, formaldehyde is one of the products of the passage of an electrical discharge through a suitable reducing gas mixture. It is not surprising, therefore, that in his early researches on the synthesis of prebiotic compounds in Earth's primordial atmosphere, S. L. Miller obtained indications of the formation of sugars. The alkalinity of the aqueous solution, caused by the presence of ammonia, presumably facilitated the conversion of formaldehyde into sugars.

Deoxyribose, the molecule of which contains an oxygen atom less than ribose, has the formula $C_5H_{10}O_4$, and so it cannot result directly from the condensation of five molecules of formaldehyde, as ribose apparently can. In 1962, however, J. Oró and A. C. Cox were able to synthesize deoxyribose from a mixture of formaldehyde or glyceraldehyde, which is $(CH_2O)_3$ and is readily produced from formaldehyde, with acetaldehyde (C_2H_4O) in an alkaline aqueous solution. As noted earlier, acetaldehyde was formed in addition to formaldehyde in Miller's experiments.

When the base used as the catalyst was calcium hydroxide (lime water), the reaction leading to the production of deoxyribose took place quite rapidly at room temperature, even in a dilute solution of the aldehydes. Other bases can be used, however, and the synthesis of deoxyribose can still occur, although the process is fairly slow, at 0° C. It has been reported that sugars undergo significant decomposition in alkaline solution, particularly in the presence of air, but Oró states that deoxyribose appears to be stable "for long periods of time" at low temperature even in an aqueous ammonia solution.

Indication of the formation of both ribose and deoxyribose was obtained by C. Ponnamperuma in 1963 in the electron irradiation of a mixture of methane, ammonia, and water vapor. It was assumed that formaldehyde was the source of the sugars and so, in extending the experiments, a very dilute solution of formaldehyde was employed as starting material. This was exposed either to ultraviolet radiation or to the gamma rays emitted by the radioactive cobalt-60. In each of these experiments, both ribose and dioxyribose were detected among the products. The sugars could not have resulted from catalysis by a base because the formaldehyde solutions were slightly acidic both before and after irradiation.

Objections have been expressed to the view that sugars were produced from formaldehyde in the primitive atmosphere: first, at moderate concentrations formaldehyde tends to form a simple polymer (trimer) paraformaldehyde which is not a sugar; and second, the formaldehyde might combine with ammonia in the atmosphere to produce the very stable compound hexamethylenetetramine. It is possible, therefore, that sugars may have been

produced from formaldehyde under such conditions that there was little free ammonia and the paraformaldehyde would have decomposed into the monomer.

According to a report made by C. Ponnamperuma and N. W. Gabel in 1967, these conditions could have existed in a hot spring in the presence of natural clay minerals. They showed that by heating a moderately dilute solution of formaldehyde in the presence of a catalyst, consisting of alumina or an aluminosilicate found in clay, a mixture of sugars, including five-carbon sugars, was obtained.

Formation of Nucleosides and Nucleotides

With the availability of a suitable pyrimidine or purine base and also ribose or deoxyribose, the next step in the production of a nucleic acid would presumably be the formation of a nucleoside; that is, a combination of ribose or deoxyribose with the base. It is expected that this would be followed by addition of a phosphate residue to the nucleoside to yield a nucleotide. Nucleic acids may be regarded as long-chain polymers of four different nucleotides.

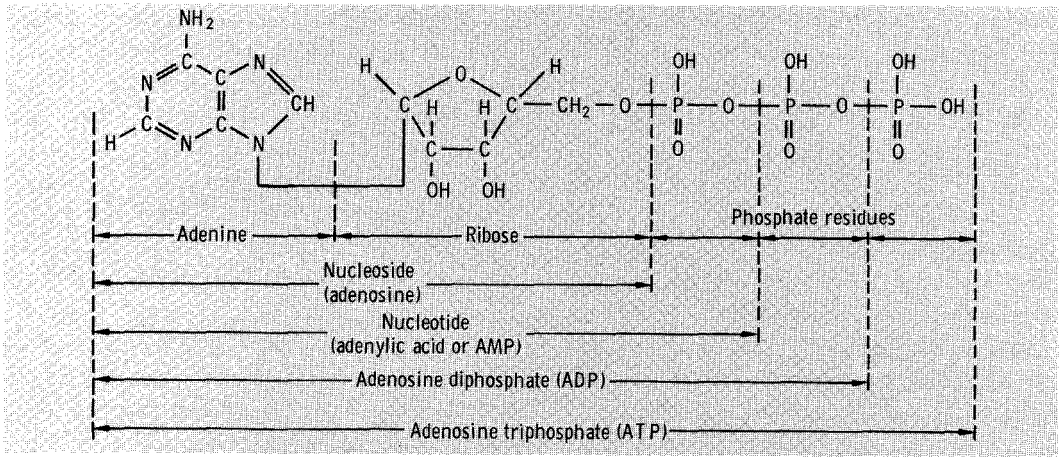
The direct synthesis of a nucleoside from its components, a nitrogenous base and a five-carbon sugar, was achieved in 1961 by G. Schramm and his associates using as a condensing agent ethyl metaphosphate, described earlier in connection with the formation of polypeptides. The pyrimidine base adenine and the sugar ribose were dissolved in formamide and the solution was heated with the ethyl metaphosphate. The nucleoside adenosine was produced in this manner. Subsequently, the corresponding deoxynucleoside was obtained from adenine and deoxyribose.

The possibility that nucleosides would be formed by the absorption of ultraviolet rays was suggested in 1961 by Carl Sagan and investigated experimentally by him together with C. Ponnamperuma and Ruth Mariner in 1963. It was found that when a dilute aqueous solution of adenine, ribose, and phosphoric acid was irradiated with ultraviolet light, adenosine was formed. The phosphoric acid in this experiment served as a condensing agent for the combination of adenine and ribose by the elimination of a molecule of water.

The adenine nucleotide (adenosine monophosphate) from adenine and ribose was obtained when ethyl metaphosphate was used as the condensing agent in place of phosphoric acid in the presence of ultraviolet light. In this work, dilute aqueous solutions were employed, in contrast to the more concentrated nonaqueous systems used by Schramm in the absence of ultraviolet radiation. Ponnamperuma and his collaborators found that when a solution containing adenine, ribose, and the ester was irradiated, the first product was adenosine. As the experiment continued, a phosphate residue became attached to the adenosine to yield the nucleotide adenosine monophosphate (AMP), also known as adenylic acid.

Under the same conditions, further phosphate groups were added to the adenoside monophosphate to give adenosine diphosphate (ADP), and then adenosine triphosphate (ATP). The general structure of these compounds is shown on page 190. Actually, AMP and ADP would have an extra hydrogen atom at the point where the additional phosphate group is attached. Among the products there was also some adenosine tetraphosphate, but, apart from the nucleotide (AMP), it is the triphosphate (ATP) that is of special interest.

In essentially all cell reactions for which energy is required, this energy is supplied by ATP. A striking fact about ATP is its uni-



versality. It appears to be present in and essential to the cells of all living organisms. The two P—O—P bonds, which serve to link each pair of phosphate groups in ATP, are said to be "energy rich." When these bonds are broken, so that the ATP is converted first into ADP and then into AMP, the energy can be made available for cellular processes requiring energy. It is thus highly significant that appreciable quantities of ATP were produced under conditions that may have simulated those existing during the prebiotic period on Earth.

An alternative synthesis of nucleotides was reported by C. Ponnamperuma and Ruth Mack in 1965. They heated each of the five nucleosides, containing one or other of the bases adenine, guanine, cytosine, thymine, and uracil, plus a ribose residue, with an inorganic phosphate at temperatures up to 160° C in a sealed tube for a few hours. In every case the corresponding monophosphate, that is, the nucleotide, was formed. The yields of the pyrimidine bases (cytosine, thymine, and uracil) were greater than for the pyridine bases. There were indications that nucleoside di- and tri-phosphates were also formed to a small extent. Although a temperature of 160° C was used in most of this work, nucleotides were produced at lower temperatures although more slowly. Thus, L. E. Orgel et al. found in 1967 that uridine, the nucleoside corresponding to uracil, could be gradually converted into the nucleotide by heating with an inorganic phosphate at 65° C.

Efficient phosphorylation of nucleosides at even lower temperatures, from 0° to 20° C, was achieved by S. W. Fox in 1967 by means of polyphosphoric acid. Both nucleotides and higher phosphates were produced in this manner. As pointed out earlier, however, it is very unlikely that such a concentrated acid would have been present on the primitive Earth.

Formation of Polynucleotides

Some indications were obtained in the experiments described above of the simultaneous production of dinucleoside monophosphates; that is, two nucleoside residues connected through one phosphate group. The formation of much more complex nucleotide polymers was reported by G. Schramm in 1963 as a result of heating nucleosides with an excess of ethyl metaphosphate in the absence of a solvent. Simple nucleotides were first formed and these then

polymerized to yield a mixture of polynucleotides of varying chain lengths. The mean molecular weight was found to be around 10 000, indicating an average chain length of roughly 30 nucleotide residues.

In natural polynucleotides (nucleic acids) a phosphate group serves to attach a specific carbon atom in one sugar residue to a different specific carbon atom in an adjacent sugar residue, as shown in the structural representation on page 167. Linkage by a phosphate group between other carbon atoms can occur, but the resulting polynucleotides are not nucleic acids.

In the polymerization of nucleotides by ethyl metaphosphate, as described by Schramm, different types of polymers were formed. A considerable proportion, but not all, of these evidently had the same structure as the nucleic acids. This was shown by their degradation into simple nucleotides by means of an enzyme which is characteristic for the breakdown of nucleic acids. Furthermore, a synthetic polynucleotide made from uracil had the same coding action for phenylalanine as did a polymer formed with the aid of an enzyme. The product thus appears to be a very simple nucleic acid possibly related to transfer RNA.

It is by no means obvious that conditions could have existed on the primitive Earth similar to those used in the laboratory to produce, in turn, nucleosides, nucleoside phosphates, and polynucleotides. Still, it is not improbable that the essential organic ingredients, the nitrogenous (pyrimidine or purine) base and the five-carbon sugar (ribose or deoxyribose), were formed from the reducing atmosphere by the action of solar ultraviolet rays, lightning discharges, heat, beta particles (electrons), etc. Phosphorus, although not one of the most abundant elements in nature, is nevertheless very widely distributed as phosphates on Earth, and so it would be available for the production of nucleotides.

So far, the polymerization of nucleotides in the laboratory, without the use of biological material, has required the use of extreme conditions. Because the polymerization can be readily performed with the aid of enzymes as catalysts, it is not impossible that other catalysts, of nonbiological origin, can be effective for converting nucleotides into polymers resembling nucleic acids. Such catalysts may have been responsible for the first polymerization of nucleotides on the primitive Earth, and they may be discovered by laboratory experiments in due course.

Coacervates and Microspheres

If, as has been assumed, life evolved on Earth from nonliving materials, there is little doubt that the later stages, at least, took place in an environment of liquid water. The presence of large proportions of water in all living organisms and its outstanding properties as a solvent are indications of the importance of this liquid for life processes. If polypeptides related to proteins and polynucleotides related to nucleic acids were formed by prebiological chemical reactions, such as those suggested in the preceding sections of this chapter, the concentrations of these substances in the oceans would probably have not been very high.

For the process of replication of deoxyribonucleic acid to take place at a significant rate, the component nucleotides must be readily available, as they are in a cell undergoing mitosis or fission. It would seem, therefore, that at some state in the development of life, a phase separation occurred. As a result, a small closed system, remotely resembling a cell, was formed. Within this closed system, with selectively permeable walls, suitable

materials might accumulate. The concentrations within the semiclosed system would be greater than in the surrounding medium and then various essential reactions could take place at a useful rate. Two possible modes of such a separation have been considered.

Many polypeptides, proteins, and other natural macromolecules, gelatine, for example, appear to dissolve when heated in water or in an aqueous medium that is slightly acid or alkaline, according to circumstances. The resulting liquid is, however, not a true solution, like that of a simple amino acid (or other simple molecule), but what is known as a colloidal solution. This situation arises because of the high molecular weight and large size of the macromolecules.

Under certain conditions, most of the colloidal polymeric material can be made to separate out from the aqueous medium in the form of very small droplets consisting of both polymer and water. The colloid-rich liquid that separates in this manner is called a coacervate. A common method for producing coacervates is to mix two colloidal solutions, one containing a macromolecular species, such as a polypeptide, with predominantly acidic properties caused by —COOH groups, and the other a polymer in which basic ($-NH_2$ or $=NH$) groups predominate.

According to the views developed by A. I. Oparin, coacervates played an essential role in the evolution of the first living organism from nonliving matter. The formation of coacervate droplets provides a means of concentrating the polymer molecules previously widely dispersed in the colloidal solution. Furthermore, the large surface area presented by the droplets to the surrounding medium facilitates rapid interchange of various soluble substances between the interior of the droplets and their surroundings.

An example of the formation of coacervate drops is seen in the microphotograph reproduced in figure 9.3. It was obtained by Oparin during the course of the polymerization of adenosine diphosphate by means of a suitable enzyme in the presence of the protein histone. After the coacervate had formed, the rate of polymerization was found to be significantly greater within the droplets than in the residual solution. This indicates that there was a higher concentration of the adenosine phosphate within the droplets than in the surrounding medium.

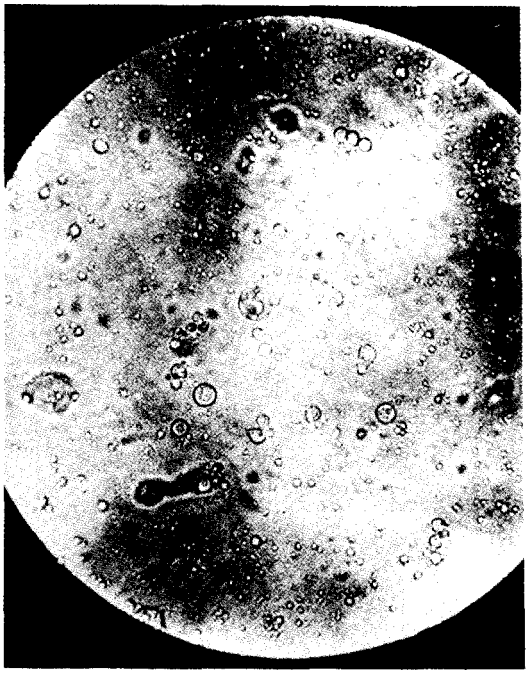


FIGURE 9.3. Photomicrograph of coacervate drops. (Photograph by A. I. Oparin; courtesy Academic Press, Inc.)

The membranes of living cells are highly selective in the manner in which they allow various dissolved substances to pass through them, in either direction. Coacervate droplets also possess a degree of selectivity with respect to their ability to take up and concen-

trate substances from outside. Oparin has found, for example, that certain dyes, such as methylene blue, as well as other compounds, including some amino acids, are extracted from the surrounding solution and accumulate within the coacervate droplets.

Another illustration of the selective behavior is provided by the coacervate formed from the colloids gum arabic and histone. Gum arabic is a complex carbohydrate, which may be thought of as a sugar polymer, whereas histone is a polypeptide. When an enzyme capable of breaking down the gum arabic into starch and the simple sugar glucose was introduced into the coacervate droplets, Oparin found that the glucose formed passed into the surrounding solution, but the starch did not. If, in addition, an enzyme which converts starch into the sugar maltose was added, the maltose also appeared outside the droplets.

It is conceivable, therefore, that processes which might not take place at all, or perhaps extremely slowly, would occur at significant rates within coacervate droplets. Upon the accumulation of reaction products inside a droplet, it would increase in size. Oparin has suggested that under the influence of external mechanical forces, such as waves, tides, or wind, the enlarged droplets would break in two, thereby providing a model for cell division.

Another possible method of realizing a closed system involves the proteinoids described earlier in connection with the synthesis of polypeptides. S. W. Fox (1959) has found that when a proteinoid is dissolved in water by heating to the boiling point, and the solution, which is colloidal, is allowed to cool, a large number of microscopic, spherical droplets, called microspheres, are formed. The individual spheres, shown in figure 9.4, have an average diameter of about 2 microns, which is approximately the same size as many spherical (coccoid) bacteria. The microspheres were also produced by R. S. Young by slowly cooling a proteinoid solution from 25° to 0° C, without intermediate heating. The spheres are usually very stable, retaining their shape for long periods of time.

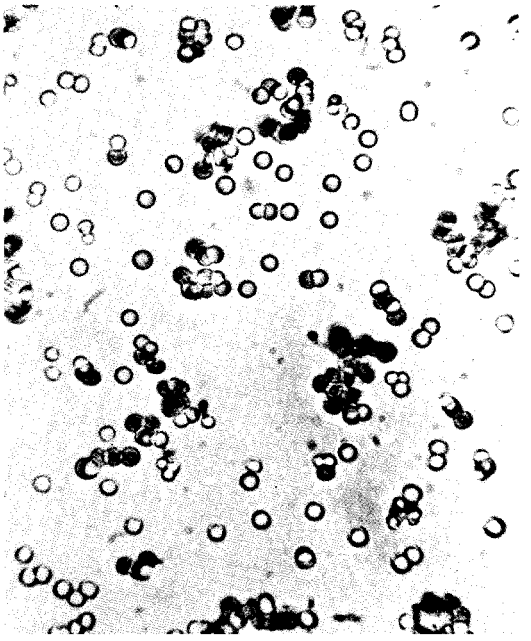


FIGURE 9.4. Microspheres of a proteinoid. (Photograph by S. W. Fox; courtesy Academic Press, Inc.)

The microspheres show several similarities to living cells and bacteria. The double boundaries seen in figure 9.5, for instance, are suggestive of a cell membrane. Moreover, like cells, the proteinoid spheres can be made to swell or shrink, due to passage of liquid in one direction or another through the surface, by placing them in aqueous solutions of different concentrations. The response is, however, less marked than it is for living cells. A decrease in the acidity of the medium in which the microspheres are suspended sometimes results in an appearance analogous to that of a cell just about to divide (fig. 9.6).

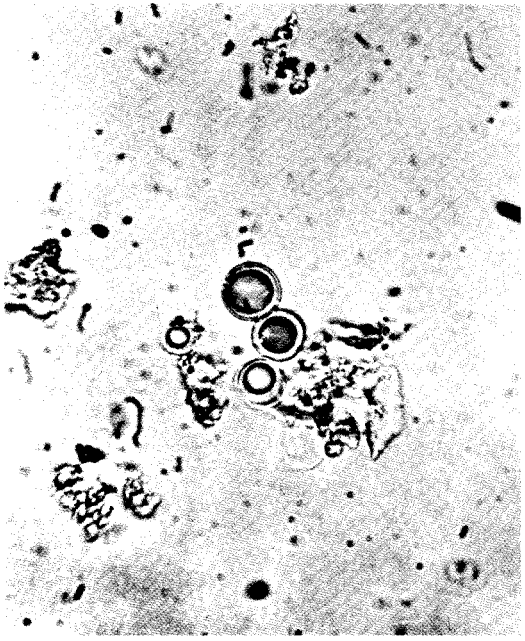


FIGURE 9.5. Proteinoid microspheres with double boundaries. (Photograph by S. W. Fox; courtesy Academic Press, Inc.)

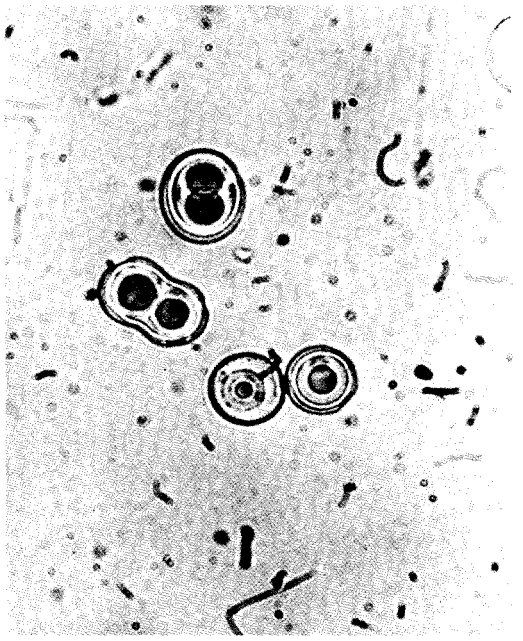


FIGURE 9.6. Twinned proteinoid microspheres obtained by decrease in acidity. (Photograph by S. W. Fox; courtesy Academic Press, Inc.)

The microspheres exhibit the Gram stain reaction which is used to classify bacteria as gram positive or gram negative. By varying the proportions of acidic and basic proteinoids used in the preparation of the microspheres, it is possible to produce either gram-positive or gram-negative states. The solubilities of the gram-positive and gram-negative microspheres in dilute alkali closely parallel those of the corresponding bacteria.

In an effort to simulate a possible mode of formation of proteinoid microspheres in nature, S. W. Fox and R. McCauley (1963) obtained a piece of lava from Hawaii. In a depression they placed a mixture of amino acids which produced a proteinoid upon heating. The piece of lava was then heated for several hours at 170° C, a temperature similar to that observed near active Hawaiian volcanoes. Upon washing with a spray of water to represent rain, a turbid liquid containing a large number of microspheres was obtained.

Some of the properties of coacervates and proteinoid microspheres are certainly interesting and perhaps suggestive. But it is of course impossible to say whether or not such systems played a role in the development of actual living cells. If primitive single-celled organisms, capable of replication, did indeed evolve from simpler systems, then it is not impossible that such systems were similar to, although perhaps not identical with, coacervates or microspheres produced in the laboratory. In any event, the purpose of the foregoing discussion is to indicate possible stages in the development of life. As to their probability, however, nothing can be said because of limitations in available knowledge.

The Origin of Optical Activity

Asymmetric compounds, such as the amino acids, except glycine, and sugars which are found in living organisms, exhibit the phenomenon of optical activity (p. 164). Such substances can exist in two forms (optical isomers) which rotate the plane of polarization of light in opposite directions. When synthesized in the laboratory, the product is invariably a racemic mixture, consisting of equal amounts of the dextro- and levo-rotatory forms. The same situation would undoubtedly have existed in the production of prebiological asymmetric compounds on the primitive Earth. The problem then is to account for the fact that, with a few minor exceptions, living organisms all contain only one of the two possible configurations of optically active compounds, namely the L-configuration of amino acids and the D-configuration of sugars.

"It is almost certain," say N. H. Horowitz and S. L. Miller, "that the presence of L-amino acids in all living organisms stems from the presence of L-amino acids in the most primitive organism from which all the others have evolved. The question is then, how the organism had obtained its optical activity." Several different suggestions have been made to account for the asymmetric synthesis, that is, the preferential formation of molecules with one type of configuration, which presumably occurred in the early development of life.

Toward the end of the 19th century, the eminent Dutch physical chemist, J. H. van't Hoff, who was one of the first to relate optical activity to the asymmetric carbon atom, suggested that circularly polarized light might have led to the appearance in nature of one isomer of an optically active pair. In circular polarization, the electric vector, instead of remaining in one plane as in plane polarized light (fig. 5.5), is like a continuously rotating radius of a circle.

Several years later, in 1934, it was shown in the laboratory that in certain photochemical reactions, that is, reactions occurring under the influence of light, there is a slight preferential formation of one of the two possible optical isomers if the light is circularly polarized. Light reflected from the surface of water is elliptically polarized, which is similar to circular polarization except that the electric vector varies continuously in magnitude instead of remaining constant. Consequently, asymmetric synthesis could conceivably have occurred to some extent in nature.

A second possibility was suggested in 1950 by the British physicist J. D. Bernal. In 1908, I. Ostromisslensky in Russia had noted that asymmetric reactions might be achieved in the laboratory by using optically active crystals as catalysts, and confirmation of this idea came from G. M. Schwab and his coworkers in Germany in the early 1930's. Quartz crystals found in nature are invariably optically active, some being dextrorotatory whereas others are levorotatory. Consequently, Bernal thought that the optical dissymmetry found in living organisms may have arisen from the synthesis of prebiotic compounds in reactions catalyzed by quartz particles. Presumably such synthesis would have had to take place entirely, or preferentially at least, on crystals having one type of activity (dextro or levo).

If laboratory experiments provide any indication, it would seem that the preference of one configuration over the other in asymmetric synthesis would have been very small. Consequently, it is not easy to understand how one structural form obtained complete dominance over the other. A different point of view, which does not postulate asymmetric

synthesis, has been favored by the American biologist George Wald, in particular. He contended in 1957 that the present situation is the result of "natural selection at the molecular level." In other words, the use of a single configuration is considered to have been advantageous to primitive organisms. This suggestion is admittedly speculative and there is no experimental evidence in its favor. Nevertheless, it is not unreasonable and is consistent with current concepts of the evolution of living species based on natural selection.

The first primitive enzyme could have been made up entirely of D-amino acids only, of L-acids only, or of some D- and some L-forms. Because of their configurations, the D-enzymes would have been catalytically active only in reactions involving D-amino acids, and the L-enzymes with L-amino acids. Such enzymes might well have been more effective than those in the third category, which would require D-amino acids at some points and L-amino acids at others. It has been suggested, too, that the mixed-configuration enzyme might have been less stable than the D-only or L-only forms. No difference would be expected, however, between the enzymes in the two latter categories.

It is possible to conceive, therefore, of an organism with two sets of enzymes, one consisting of D-amino acids only and the other of L-amino acids only. To produce two separate systems having the same functions would probably have been a drawback to a primitive organism. The tendency would be, therefore, for the organism to discard one or the other of the two equivalent possible enzyme systems.

If life developed just once on Earth, the organism might well have used only L-amino acids, as is the case at present. On the other hand, if life originated at several places at different times, it is expected that some forms would be based entirely on D-amino acids, whereas others would consist of L-configurations only. As the result of a change (or mutation), however, possibly caused by the action of ultraviolet radiation, one particular variety, say the L-form, might have developed a selective advantage over the other. The organisms using the L-enzymes would then increase much more rapidly, and the D-system would tend to disappear. The foregoing discussion has referred in particular to amino acids, but a similar situation could have arisen with the sugars, leading to the exclusive use of the D-type configurations in living organisms.

THE CARBONACEOUS CHONDRITES

Types of Meteorites

Information that may have a bearing on the origin of life, or, at least, on the formation in nature of various materials from which life might have developed, has been obtained from a study of certain types of meteorites. A meteorite is an extraterrestrial object, commonly associated with a fireball, that survives passage through the atmosphere and reaches Earth's surface. Up to the present, about 1500 individual meteorites have been identified. Of these, roughly half have actually been observed soon after reaching Earth, and are called falls. The remaining half or so, referred to as finds, have been discovered independently.

The great majority of meteorites are of two types: the stony meteorites (stones) which, like terrestrial stones, consist mainly of silicates of calcium, magnesium, iron, aluminum, etc.; and the iron meteorites (irons), which are made up largely of iron with various proportions of nickel and small quantities of other metallic and nonmetallic elements. The stones constitute some 61 percent of known

meteorites and the irons represent some 35 percent. The remaining 4 percent or so are the stony irons which are intermediate in character.

Most (over 90 percent) of the stony meteorites belong in the category of chondrites. They have been given this name because they have small spheroidal inclusions, a millimeter or two in diameter, called chondrules (from the Greek word *chondros,* meaning grain of wheat). Thus somewhat more than half, roughly 800, of the identified meteorites are probably chondrites.

Among the chondrites are about 20 specimens of falls which are exceptional in the respect that they are black in color and friable in nature, so that they crumble readily. These chondrites contain roughly 2 to 4 percent of carbon in the form of various compounds, and they are known as carbonaceous chondrites. Although such meteorites constitute slightly more than 2 percent of the total number of chondrites, their actual frequency is undoubtedly larger, perhaps 5 percent. All the known carbonaceous chondrites were observed to fall to Earth, but there certainly were others that fell without being seen. Because they are so friable, carbonaceous chondrites that are not picked up soon after reaching Earth disintegrate in a short time and are thus lost. Others may well disintegrate during their passage through the atmosphere and form small undetectable pieces. The largest reported carbonaceous meteorite weighed about 100 pounds, compared with over a ton for the largest noncarbonaceous chondrite.

Two main categories of carbonaceous chondrites have been recognized, and there may be a third category containing somewhat less carbon, but this is not of immediate importance. The chief difference between the two established classes of carbonaceous meteorites is that one has somewhat larger proportions of carbon, water, and sulfur, but a lower specific gravity than the other. The approximate average values of these properties are given below. For comparison, it may be noted that noncarbonaceous chondrites contain, on the average, about 0.16 percent of carbon and 1.82 percent of sulfur. The most important component, roughly half, of all carbonaceous chondrites is magnesium silicate, with small amounts of aluminum, calcium, and other silicates; in addition, there is around 20 percent of iron oxide and sulfide.

	Carbon	Water	Sulfur	Specific gravity
Type I, percent..	3.5	20	6.2	2.2
Type II, percent.	2.5	13	3.3	2.5–2.9

Compounds in Carbonaceous Chondrites

The first recorded carbonaceous chondrite was seen to fall in 1806 near the town of Alais in the south of France. A sample was sent to the eminent Swedish chemist J. J. Berzelius, and in 1834 he reported that analysis indicated the presence of material resembling the humus, a product of vegetable decay, in terrestrial soils. Berzelius wondered whether this might indicate that living organisms existed, or had existed, on extraterrestrial bodies, but apparently he concluded that such was not the case.

In 1838, a second carbonaceous meteorite was recorded in the vicinity of the Cold Bokkeveld Mountains in South Africa. From this, the German chemist Friedrich Wöhler extracted a liquid which he described in 1850 as having a "strong bituminous odor." Wöhler later investigated another carbonaceous chondrite which had fallen at Kaba in Hungary in 1857, and he concluded that

the carbon compounds in these meteorites were of biological origin. It is of interest that Wöhler came to this conclusion in spite of the fact that his synthesis of urea from inorganic materials in 1828 had cast the first doubt on the view, universally accepted at the time, that the organic compounds could be produced only by living organisms.

One of the best known of the (type I) carbonaceous chondrites fell near the village of Orgueil in southern France in 1864. It was seen to break up as it neared Earth and about 20 pieces of various sizes, most of them 3 inches or less in diameter, were recovered. From an analysis of one specimen, made soon after its fall, the French chemist S. Cloëz concluded that the organic material present was similar to that in peat and lignite. On Earth, these substances are formed from vegetable matter. An examination of the Orgueil meteorite was also made by the well-known French chemist M. P. E. Berthelot and he reported in 1868 that saturated hydrocarbons, similar to those in petroleum, were present. Berthelot thought they were of nonbiological origin because he considered that terrestrial petroleum is derived from inorganic sources. There is evidence at present that much petroleum is of biological origin. In more recent work, hydrocarbons of several different types have been found in carbonaceous meteorites.

For almost 90 years, the carbonaceous chondrites appear to have attracted little attention from the standpoint of their chemical composition. But several studies were made of the possible presence in them of the remains of living organisms. Then in 1953, G. Mueller in England published the results of a detailed analysis of the elements in material extracted from the Cold Bokkeveld (type II) meteorite by means of various solvents including water. The dried extraction product was a brown, amorphous, sticky solid with a faint aromatic odor. Apart from 15 to 20 percent of mineral ash and some free, elemental sulfur, the compounds consisted mainly of the elements carbon and oxygen, with a few percent each of hydrogen, nitrogen, sulfur, and chlorine. This extract, which constituted only about one-fifth of the carbonaceous material in the meteorite, is evidently a very complex mixture of compounds, the exact nature of which is still in doubt.

Compounds of Possible Biological Significance

The current interest in the composition and origin of carbonaceous meteorites corresponds with the development of the modern space age and the possibility of finding extraterrestrial life, especially on Mars. In 1959, Melvin Calvin of the University of California, Berkeley, who was interested in the problem of the development of life, obtained a sample of a carbonaceous chondrite (type II) that had fallen near Murray, Ky., in 1950. A series of analytical procedures led M. Calvin and Susan K. Vaughn to conclude that the meteorite contained several complex molecules, including hydrocarbons with chains of at least 15 carbon atoms. More significant, however, was an indication that the ultraviolet spectrum of some of the material extracted from the meteorite by acidified water resembled that of cytosine, one of the pyrimidine bases present in the DNA molecule. Later investigators, however, suggested that the substance may have been an impurity introduced inadvertently during the course of the experimental work (p. 201).

Although many scientists of the 19th century were of the opinion that the carbon compounds in carbonaceous meteorites originated from living matter, this view had fallen into disrepute. Calvin, in fact, concluded that the organic compounds he had detected

were not of biological origin. He regarded their presence simply as evidence that substances from which life could have evolved existed on extraterrestrial bodies and that such prebiotic materials would presumably have also been present on Earth.

Hydrocarbons in Carbonaceous Meteorites

Early in 1961, however, there occurred a controversial development in the situation concerning the significance of the organic compounds in carbonaceous chondrites. This arose from a report made by B. Nagy, W. G. Meinschein, and D. J. Hennessy in the United States of their work on the Orgueil meteorite. The sample was heated and the volatile materials, which vaporized at various temperatures, were collected and subjected to analysis with a mass spectrometer. The products were mainly saturated hydrocarbons, containing up to 29 carbon atoms, and the relative amounts of different types were said to be similar to the proportions of such hydrocarbons found in marine sedimentary deposits on Earth and in butter. Because these terrestrial sedimentary deposits and butter are undoubtedly of biological origin, the conclusion drawn was that "the composition of the hydrocarbons in the Orgueil meteorite provides evidence of biological activity." In other words, the implication was that the carbonaceous material, or at least part, was produced by extraterrestrial living organisms.[1]

In extension of the work described above, the same three scientists in 1963 extracted hydrocarbons from samples of both the Orgueil and Murray carbonaceous chondrites, by using a mixture of benzene and methyl alcohol as solvent. The extract was separated into four fractions by means of a well-established chromatographic procedure. The fractions so obtained were examined by infrared, ultraviolet, and mass spectrometry. Again, the results indicated the presence of hydrocarbons with a marked similarity to those derived from terrestrial marine sediments.

Several objections have been raised both to the experimental results just outlined and to the conclusions drawn from them. Edward Anders in the United States, for example, suggested in 1962 that hydrocarbons resembling those present in sedimentary deposits may have been formed by purely chemical reactions in the solar nebula from which the planets developed. In this event, the hydrocarbons, and presumably other carbon compounds present in the meteorites, would have been incorporated into Earth, as well as into other bodies in the solar system. According to Anders, when life developed on Earth, those organisms which could utilize the available hydrocarbons, and subsequently synthesize them, would survive. This would consequently account for the similarities between the hydrocarbons in carbonaceous chondrites and those in some sedimentary deposits which contain the remains of living organisms.

Support for the foregoing view appears from the work reported by M. H. Studier, R. Hayatsu, and E. Anders in 1965. A mixture of carbon monoxide and hydrogen gases, such as might have existed in the primordial solar nebula, was heated in the presence of a catalyst made from particles of stony and iron meteorites. Particles of the same composition would presumably have been present in the primitive solar system. A rapid reaction occurred between the gases, and among the products were many hydrocarbons similar to those found in carbonaceous chondrites. In

[1] Not all the hydrocarbon (petroleum) oils in nature are necessarily of biogenic origin. Some, such as those in sedimentary deposits, probably are, but there may be others that are of inorganic origin.

fact, the distribution of hydrocarbons formed in the chemical reaction between carbon monoxide and hydrogen catalyzed by meteorite particles followed by subsequent heating was said to match the distribution in chondrite hydrocarbons in all important respects.

Nearly all carbonaceous chondrites have been found to contain small proportions of two long-chain hydrocarbons identified as pristane ($C_{19}H_{40}$) and the chemically analogous (homolog) phytane ($C_{20}H_{42}$). The latter is of special interest because it is closely related chemically to part (phytol) of the molecule of chlorophyll, the green coloring matter of plants. The two hydrocarbons are also present in terrestrial petroleums that are undoubtedly of biological origin. Hence, it appeared, at first sight, that the presence of pristane and phytane in the carbonaceous chondrites was a strong argument for the biogenic origin of all the hydrocarbons in these meteorites. The identification of pristane has been claimed, however, among the products obtained by Studier and his associates from the reaction of carbon monoxide and hydrogen. The presence of pristane and phytane in the carbonaceous meteorites is thus not decisive in determining the origin of the carbon compounds they contain.

It is noteworthy that the mechanism suggested above for the purely chemical formation of hydrocarbons in the solar nebula is strikingly similar to the Fischer-Tropsch process used in industry. In this process for the production of hydrocarbons and other organic compounds, a mixture of carbon monoxide and hydrogen is passed over a heated catalyst, usually consisting of iron, iron oxide, and iron carbide. W. G. Meinschein and his coworkers had reported that the infrared spectrum of Fischer-Tropsch oil was distinctly different from that of the meteorite hydrocarbons. It must be remembered, however, that the products of the Fischer-Tropsch process are dependent upon the nature of the catalyst and the temperature of interaction.

Another argument against the biological origin of the hydrocarbons in chondrites was presented by J. Oró and his associates in 1966. They made an extensive study of the hydrocarbons extracted from samples of several carbonaceous chondrites, and although the results confirmed some of those obtained by B. Nagy and his colleagues, there was one important difference. The latter had reported that the proportion of hydrocarbons with odd numbers of carbon atoms was enhanced in meteorites, a situation thought to be typical of biological material. Oró found, however, that the relatively larger amounts of odd numbers of carbon atoms occurred only in a few samples of the Orgueil chondrite, but not in other carbonaceous meteorites including several Orgueil specimens. It is conceivable that the different compositions were caused by contamination of some pieces by recent terrestrial material.

The possibility of contamination of the samples of meteorites has been mentioned by several scientists. A certain degree of contamination has undoubtedly resulted from repeated handling, from the soil upon which the meteorite fell, from matter taken in from the atmosphere by the many pores in carbonaceous chondrites, from terrestrial microorganisms entering the specimens during storage in museums, and in other ways. For example, in 1966, A. A. Imshenetsky and his associates reported that sterilized samples of meteorite pieces became severely contaminated internally with micro-organisms when exposed to moist soil, although on dry sand the contamination was only on the surface.

Other Carbon Compounds in Meteorites

The hydrocarbon constituents of carbonaceous chondrites have attracted more interest than the other organic compounds because they are readily removed from a meteorite specimen by volatilization or by extraction with a solvent. Furthermore, they can be analyzed by techniques that have been developed in petroleum chemistry. Nevertheless, the identified hydrocarbons represent only a small fraction, little more than 1 percent, of the total carbonaceous matter in the meteorites. A few other types of carbon compounds have been characterized. They are aromatic (benzenoid) acids and phenols (3.3 percent), fatty acids (0.75 percent), and nitrogenous compounds (about 0.25 percent).

Approximately 95 percent of the organic compounds in carbonaceous chondrites is still unidentified. Most of this is an insoluble black material consisting of various macromolecular species, including complex hydrocarbons. Such material is extremely difficult to characterize except in very general terms. Incidentally, much of the carbonaceous matter in terrestrial sediments of biological origin has also not yet been identified.

Of the nonhydrocarbon material in carbonaceous chondrites, reference will be made here only to those compounds which may possibly have some bearing on the origin of life; namely, nitrogenous compounds and fatty acids. It will be recalled that in 1960, M. Calvin reported the extraction from the Murray meteorite of a substance with an ultraviolet spectrum resembling that of the pyrimidine base cystosine. In the following year, Michael H. Briggs in New Zealand examined a sample of the carbonaceous chondrite which had fallen at Mokoia in that country in 1908. He stated that he had obtained evidence of the presence of purine bases, but could not identify them. The possible occurrence of pyrimidines and purines is of special interest because they are essential components of the nucleic acids.

Because of the suggestion that the spectroscopic patterns attributed to these nitrogenous bases may have been caused by impurities derived from ion-exchange resins used in the extraction process, Ryoichi Hayatsu, at the University of Chicago, attempted in 1965 to isolate the bases from a carbonaceous chondrite without using such resins. A powdered sample of the Orgueil meteorite, which had been previously treated with water and several organic solvents and could thus be presumed to be free from contaminants, was heated for several hours with hydrochloric acid in a sealed tube. From the solution obtained in this manner, a number of fractions were obtained by extraction with solvents.

The purines adenine and guanine, present in nucleic acids, were identified among the products by their infrared and ultraviolet spectra. Other organic substances found were melamine and the related compound ammeline, neither of which has any known biological significance. It may be noted, however, that melamine, which has the formula $C_3H_6N_6$ or $(CH_2N_2)_3$, can be obtained by the polymerization of cyanamide, formula $CN \cdot NH_2$ or CH_2N_2. The latter, as mentioned earlier, could well have been formed under the conditions existing on the primordial Earth.

Both M. Calvin and M. H. Briggs had tested for amino acids in their extracts from carbonaceous chondrites, but the results were negative. A similar conclusion concerning the apparent absence of amino acids was reached by P. B. Hamilton in 1964. Some other scientists claimed to have obtained indications of the presence of indigenous amino acids, not caused by contamination after the meteorite had fallen to Earth. On the whole, however, it would seem that the results of the

tests for free amino acids in carbonaceous chondrites are not decisive. It appears that if any of these compounds are present in the uncontaminated meteorite, the amount must be quite small; it would not account for more than about 1 percent of the total nitrogen present.

Another type of nitrogenous compound of biological interest was found in samples of the Orgueil meteorite by G. W. Hodgson and B. L. Baker in 1964. This substance is vanadyl porphyrin. The portion of the hemoglobin molecule responsible for the red color of blood is an iron porphyrin, and chlorophyll in green plants contains a residue of a magnesium compound of chlorin, which is closely related to porphyrin. The infrared spectrum of vanadyl porphyrin extracted from the Orgueil chondrite was found to resemble that from a German shale which is a typical source of ancient biological porphyrins. According to Hodgson and Baker, the results suggest "a strong possibility of biogenic agency in the origin of the organic matter of the meteorite."

In 1967, however, C. Ponnamperuma and G. W. Hodgson claimed to have obtained evidence of the formation of porphyrins by passage of an electric arc through a simulated primordial atmosphere of ammonia, carbon dioxide, and water vapor. The vanadyl porphyrin found in the Orgueil meteorite could thus have been of prebiological origin.

It should be mentioned that the organic (carbon) compounds, which constitute roughly 5 to 10 percent of a carbonaceous chondrite, contain from 2 to about 5 percent of the element nitrogen. But the pyrimidines, purines, porphyrins, and, possibly, amino acids isolated from the meteorites can account for only a small fraction of this amount of nitrogen. It would be of great interest to know the identity of the nitrogenous compounds in which most of the nitrogen occurs in carbonaceous meteorites.

The so-called fatty acids, each of which contains a long chain of carbon atoms and a terminal acidic (—COOH) group, are found in soils, peats, lignite, and some kinds of petroleum. The fats in living organisms are compounds (esters) of fatty acids with glycerol. Fatty acids and their derivatives found on Earth are invariably of biological origin. It may be significant, therefore, that in 1963, B. Nagy and M. C. Bitz were able to extract several fatty acids from the Orgueil meteorite. The infrared spectrum and composition of the mixed acids closely resembled those of fatty acids in a sample of soil. Specimens of noncarbonaceous chondrites treated in the same manner gave no such spectrum. These results were offered as further support of the theory that the organic compounds in carbonaceous chondrites were derived from living systems.

Optical Activity of Carbonaceous Chondrites

The phenomenon of optical activity is characteristic of compounds of biogenic origin, whereas the same substances from nonbiological sources are optically inactive (racemic) mixtures of dextro- and levorotatory forms. A test for optical activity in organic materials extracted from carbonaceous meteorites might therefore throw light on their origin. In 1964, B. Nagy and his coworkers reported measurable optical activity in extracts from samples of the Orgueil meteorite. The solutions were levorotatory, whereas the most probable contaminants, such as dust from the museum in which the specimens had been stored, pollen grains, and washings from human hands, were all dextrorotatory. Extracts from noncarbonaceous chondrites gave no optical activity. These results seem to indicate that the carbon compounds in the meteorites had a biological

origin, but the source was not terrestrial.

The reliability of the results obtained in the tests for optical activity was questioned by R. Hayatsu in 1965. The latter has stated that the solutions obtained by Nagy were colored, so that they absorbed visible light strongly, and also that they contained very small (colloidal) particles which caused light to be scattered. Under these conditions, the instrument (polarimeter) used to observe optical rotation is known to give false readings. When precautions were taken to obtain extracts that were almost colorless and free from light-scattering particles, Hayatsu claimed the solutions exhibited no optical activity. Further studies by both Nagy and Hayatsu have caused neither to change his views, and so the situation remains unresolved.

It has been pointed out by C. Sagan that there might be another explanation for the levorotatory character of the extracts of the Orgueil meteorite, assuming that they do indeed possess optical activity. Suppose the compounds present in the carbonaceous chondrites do not have a biogenic origin. They would then consist of racemic mixtures when the meteorite fell to Earth. In the course of many decades, in which the specimens were stored in museums, they could have become contaminated with terrestrial organisms. These organisms would then have consumed the dextrorotatory compounds present, thereby converting them into optically inactive waste products. Consequently, the levorotatory components of the racemic mixtures would remain in the meteorite and would account for the claimed optical activity of the extracts.

Organized Elements in Meteorites

A new and highly controversial aspect of the problem of the carbonaceous chondrites appeared in 1961 when George Claus, of the New York University Medical Center, and B. Nagy observed numerous microscopic particles, which they described as "resembling fossil algae," in the Orgueil and Ivuna (Tanganyika) carbonaceous meteorites. Later, similar particles were reported in other carbonaceous chondrites. Five types of these organized elements, as they were called, were classified according to their shape, size, and surface structure as seen under the microscope. The great majority of the organized elements were spherical with diameters ranging from 4 to about 20 microns. Others, less common, were shield shaped or cylindrical, and a very small proportion were hexagonal.

From their appearance and other properties, Claus and Nagy concluded that the organized elements were the fossilized remains of living organisms. Several other observers, notably F. L. Staplin of Imperial Oil, Ltd., Canada, and B. V. Timofeyev, a Russian authority on pollens, have supported this point of view.

There is no doubt, from the many reports on the subject, that the carbonaceous meteorites do contain small, distinguishable microscopic particles with an average diameter around 10 microns (fig. 9.7), but opinion concerning their nature is divided. The main arguments for considering the so-called organized elements to be inorganic material, rather than fossilized life forms, have been presented by F. W. Fitch, a pathologist at the University of Chicago, in conjunction with E. Anders. In a detailed discussion of the situation, published in 1963, they state that the organized elements fall into two main classes. The first class, representing the great majority, consists of brownish-yellow, somewhat irregular, roughly spherical granules, whereas the second class is made up of particles with a much more complex structure.

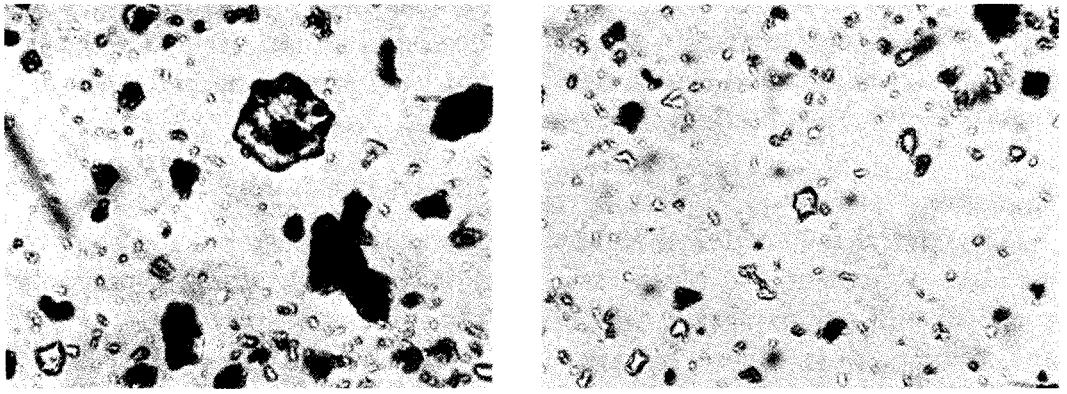


FIGURE 9.7. Microscopic structures in carbonaceous chondrites: left, from South Africa; right, from the United States. (Courtesy Jet Propulsion Laboratory, California Institute of Technology.)

According to Fitch and Anders, the organized elements in the first category do not fluoresce in ultraviolet light and they do not take biological stains in a manner that permits them to be distinguished from irregular silica grains in the meteorites. Furthermore, these particles disappear after treatment with acid. They seem to be inorganic in composition and origin, and it was concluded that the objects in the first category were probably microscopic mineral particles that were embedded in the carbonaceous chondrite. The less common organized elements, with the more complex structure, were said to be un-

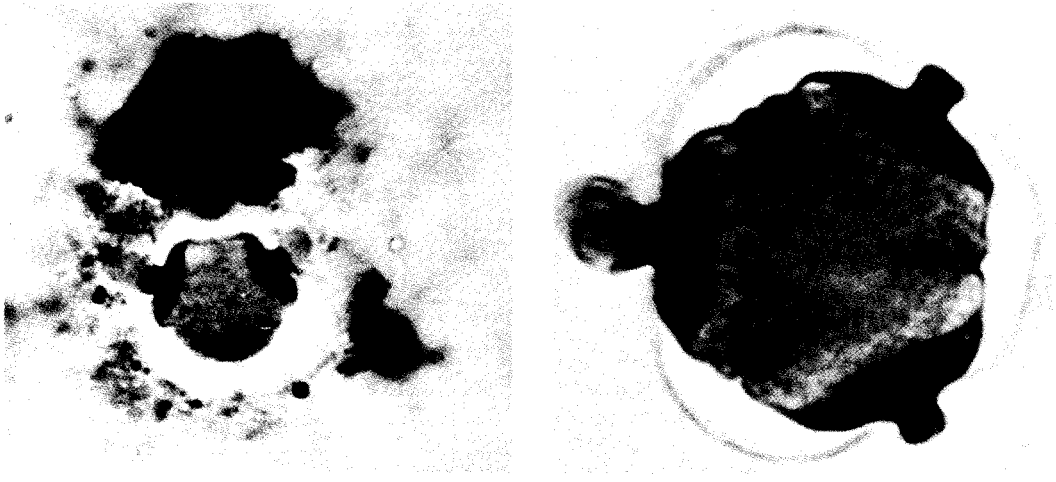


FIGURE 9.8. Organized element from carbonaceous chondrite (left) and ragweed pollen grain (right), both stained by the Gridley technique. (Photomicrographs by E. Anders, F. Fitch, and B. Nagy; from *Science*, vol. 138, pp. 1392–9, Dec. 28, 1962. Copyright 1962 by the American Association for the Advancement of Science.)

observable in thin sections and to show a resemblance to common airborne contaminants.

One of the most striking of the organized elements observed by Claus and Nagy was of the rare hexagonal form. A microphotograph obtained after staining by the Gridley technique, which is used in the study of fungi, is shown in figure 9.8. Anders and Fitch were able to secure an almost identical structure by staining ragweed pollen using the Gridley method. The staining procedure evidently distorted the pollen grains into a hexagonal shape. It appears, therefore, that the uncommon hexagonal organized elements in carbonaceous chondrites may have resulted from contamination by terrestrial pollen grains.

Another experimental approach to determination of the identity of the organized elements was made by B. Nagy and his associates in 1963. Examination with the aid of an electron microprobe, which is an extremely sensitive method of chemical analysis, indicated that some of the organized elements are impregnated with iron compounds, and some with nickel, whereas others contain silicate. After removing the mineral constituents by means of acids, there often remained a residue with the same shape as the original particle, but it contained only light elements believed to be organic in nature. It appeared, therefore, that the organized elements were not pollen grains, but they might have been living organisms that had become fossilized by minerals, in much the same way as microfossils are formed on Earth. Anders contends, on the other hand, that the particles studied were merely mineral grains.

General Conclusions

There are two important questions relating to the carbonaceous chondrites: Are the carbon compounds present of biogenic origin? Are the organized elements fossilized remains of living organisms? Unfortunately, it is at present impossible to give a definite answer to either of these questions. If the answer to one or the other were affirmative, there would be an implication, although not necessarily definite proof, as will be seen shortly, of the occurrence of extraterrestrial life, at least in the distant past. Even if both questions are answered in the negative, the fact that certain meteorites contain several compounds of biological interest, as they undoubtedly do, may be of great significance.

To develop this matter further, it is necessary to consider the origin of meteorites. Here, again, the situation is equivocal. It should be noted at the outset that meteorites probably have a different origin than do the very small meteoritic particles which produce the familiar streaks of light in the sky known as shooting stars. The great majority of the latter are undoubtedly the remains of comets, and meteor showers occur where the orbit of Earth intersects the orbit of the cometary debris (p. 147). The meteorites, which are observed to fall or are found on Earth's surface, are very much larger than the particles that produce meteors, and they are usually associated with violent disturbances such as fireballs or bolides.

There is only one case on record for which there is sufficient evidence, obtained from photographs, to permit the orbit of a meteorite to be determined with reasonable accuracy. The results indicate an asteroid-like orbit lying between Jupiter and Mars at aphelion, that is, when most distant from the Sun, and between Earth and Venus at perihelion, when closest to the Sun. This is in agreement with the widely accepted view that the meteorites originate in the asteroidal belt, much of which lies between Jupiter and Mars.

One theory of the origin of meteorites is that they were produced by the disintegration of a single planetary body with a size intermediate between that of Earth and that of the Moon. This body would possibly represent the "missing" planet between Mars and Jupiter. The breakup of such a planet is assumed to have produced the asteroids and smaller masses which appear on Earth as meteorites. Normally, these bodies remain in the asteroidal belt, but occasional collisions or close approaches result in changes in the orbits of the smaller masses so that they reach Earth. Apart from other difficulties, a serious objection to this theory is that there is no reasonable way to account for a body of planetary dimensions disintegrating into a large number of smaller masses while retaining the characteristic structures observed in meteorites, especially in the stones.

Another possibility is that the matter in the asteroidal belt is the remains of the primordial solar nebula from which the planetary system was formed. It is supposed that the effect of the enormous mass of Jupiter prevented the aggregation of the material into a planet. This theory also encounters difficulties, especially in explaining the heating that was undoubtedly responsible for some aspects of meteorite structure.

Disregarding the details of their origin, suppose that meteorites, including the carbonaceous chondrites, are of extraterrestrial origin. The presence of complex organic compounds in these bodies would then imply that such compounds were formed either on another planet, not Earth, or in the solar nebula from which the planets evolved. If the carbon compounds were derived from living organisms, as some believe, then life must have existed elsewhere than on Earth.

It should be pointed out, however, that if meteorites come from the asteroidal belt, the physical conditions, particularly the temperature, would probably not have been conducive to the development of life. Nevertheless, even if the compounds in the carbonaceous chondrites were derived from inorganic, nonliving sources, it can still be argued that materials would be available on extraterrestrial bodies, and presumably also on Earth, from which life might conceivably have originated if the conditions were suitable. If this were so, the carbon compounds present in carbonaceous meteorites would be similar to some, at least, of the prebiotic material on Earth. It is expected that the latter would have included amino acids and the absence of significant quantities of these substances in the chondrites might possibly be accounted for by decomposition over a long period of years.

Reference must be made, finally, to another point of view concerning the origin of meteorites. Some scientists, H. C. Urey in particular, contend that meteorites consist of material that came initially from the lunar surface.[2] One theory of the origin of the Moon is that it was part of Earth and separated from it, possibly at the time when Earth's metallic core was formed. If meteorites were derived from the Moon's surface and the Moon was once part of Earth, then the compounds in carbonaceous chondrites might have been formed on Earth. Even if the Moon was never part of Earth but was captured by it, in accordance with an alternative theory, it is still possible, as Urey has suggested, that material from Earth may have been thrown to great distances as a result of the disturbance that occurred at the time. Such material may have become the basis for meteorites that eventually returned to Earth.

[2] Tests made in 1967 and 1968 by instruments on the Surveyor spacecraft indicate that the present composition of the lunar surface is different from that of chondritic meteorites.

In any event, if the foregoing arguments are correct, the organic compounds in carbonaceous chondrites must have been formed more than 4 billion years ago when the Moon separated from or was captured by Earth. It is probable that there were no living organisms on Earth at that time, and so the carbon compounds in meteorites would resemble, or be part of, the prebiological materials on Earth. The general conclusion to be drawn is that there were organic compounds on Earth before living organisms appeared and consequently life might possibly have evolved from them.

FOR FURTHER INFORMATION

ABELSON, P. H.: Chemical Events in the Primitive Earth. Proc. Natl. Acad. Sci., vol. 55, 1966, p. 1365.

BRIGGS, M. H.; AND MAMIKUNIAN, G.: Organic Constituents of the Carbonaceous Chondrites. Space Sci. Rev., vol. 1, 1962–1963, p. 647.

CLOUD, P. E.: Atmospheric and Hydrospheric Evolution on the Primitive Earth. Science, vol. 160, 1965, p. 729.

ECHLIN, P.: Origins of Photosynthesis. Science J., vol. 2, no. 4, 1966, p. 42.

ECK, R. V.; ET AL.: Thermodynamic Equilibrium and the Inorganic Origin of Organic Components. Science, vol. 153, 1966, p. 628.

FERRIS, J. P.; SANCHEZ, R. A.; AND ORGEL, L. E.: Studies in Prebiotic Synthesis: III, Synthesis of Pyrimidine from Cyanoacetylene and Cyanate. J. Mol. Biol., vol. 33, 1968, p. 693.

Fox, S. W.: The Origin of Life. Nature, vol. 205, 1965, p. 328.

Fox, S. W., ED.: The Origins of Prebiological Systems. Academic Press, 1965.

MAMIKUNIAN, G.; AND BRIGGS, M. H., EDS.: Current Aspects of Exobiology. Pergamon Press, 1965.

MASON, B.: Organic Matter From Space. Sci. Am., vol. 208, no. 3, 1963, p. 43.

NAGY, B.: Carbonaceous Meteorites. Endeavour, vol. 27, no. 101, May 1968, p. 81.

Organic Matter in Carbonaceous Chondrites. Science, vol. 152, 1966, pp. 102–107.

PITTENDRIGH, C. S.; VISHNIAC, W.; AND PEARMAN, J. P. T., EDS: Biology and the Exploration of Mars. Natl. Acad. Sci.-Natl. Res. Council Publ. 1296, 1966.

PONNAMPERUMA, C.; AND GABEL, N. W.: Current Status of Chemical Studies on the Origin of Life. Space Life Sci., vol. 1, 1968, p. 64.

SAGAN, C.: Radiation and the Origin of the Gene. Evolution, vol. 11, 1957, p. 40.

SHKLOVSKII, I. S.; AND SAGAN, C.: Intelligent Life in the Universe. Holden-Day, 1966.

SHNEOUR, E. A.; AND OTTESEN, E. A., COMPS.: Extraterrestrial Life: An Anthology and Bibliography. Natl. Acad. Sci.-Natl. Res. Council Publ. 1296A, 1966.

SULLIVAN, W.: We Are Not Alone. McGraw-Hill, 1964.

Symposium on the Origin of the Earth's Atmosphere. Proc. Natl. Acad. Sci., vol. 53, 1965, pp. 1169–1226.

UREY, H. C.: Biological Material in Meteorites: A Review. Science, vol. 151, 1966, p. 157.

X

The Possibility of Life on Mars

EXOBIOLOGY OF MARS

Introduction

THE NEWLY DEVELOPING SUBJECT of exobiology is largely concerned with the study of various aspects of the existence, characteristics, and evolution of life on bodies other than Earth. The problem of immediate interest is, of course, related to the occurrence of life on Mars. In this connection, it is appropriate to ask: What are the prospects of finding living organisms on Mars?

The answer of the majority of those who have given careful consideration to this question can be summarized in the statement made by N. H. Horowitz in 1965. "Nothing we have learned about Mars," he wrote, ". . . excludes it as a possible abode of life." As to the form this life may take, nothing is known, but in a private communication to the present writer in 1968, C. Sagan went as far as to say: "I believe it is probably premature to exclude 'advanced' forms of life on Mars at our present state of ignorance."

There are several conceivable situations in connection with life on Mars. One extreme is the possibility that there never has been and never will be any indigeneous life forms, whereas the other extreme is that advanced and complex organisms now exist on the planet. The truth may well lie somewhere in between. Life may have developed in the past in more favorable circumstances but may have become extinct as a result of changes in the environment. On the other hand, more complex forms of life may have died out, leaving simpler organisms that have become adapted to the present Martian conditions.

If there is now no life on Mars, regardless of whether or not it has existed in the past, could life develop in the future? In view of the physical characteristics of the planet, this would not appear to be probable although it is by no-means impossible. Even if indigenous life forms could not arise, there are reasons for believing that some terrestrial organisms would probably survive if introduced on Mars. This raises a highly interesting question: Could man live on Mars? Any attempts to supply an answer are highly speculative at present, but it is a matter that should be exercising the thoughts of scientists, and so it will be taken up briefly at the end of this chapter.

Prebiotic Compounds on Mars

If the planet Mars had a primitive atmosphere with reducing properties, consisting of methane, ammonia, carbon monoxide (or related gases), and water vapor, then it may be supposed that the same prebiological (or prebiotic) materials would have been produced on Mars as are believed to have formed on Earth. In other words, the arguments presented in the preceding chapter would apply equally to both Mars and Earth.

Some scientists have asserted that, because the gravitational force near the surface of Mars is only about 40 percent of that on Earth, the light hydrogen gas would have escaped fairly rapidly from Mars. Consequently the reducing atmosphere, if it ever existed, would not have lasted long enough to permit the formation of the organic (carbon) compounds from which life might evolve. This is not necessarily the case. The escape of gases from a planetary atmosphere depends not only on the gravitational force but also on the temperature of the exosphere. If this temperature has always been as low as indicated by some of the Mariner IV measurements made in 1965, then hydrogen molecules would not have escaped, although atomic hydrogen would have, during the history of Mars.

Another argument is based on the presence of complex organic compounds in the carbonaceous chondrites. If, as appears probable, these meteorites originated in the asteroidal belt, which lies largely between Mars and Jupiter, then such compounds may also be expected on Mars. On the whole, therefore, there are no convincing reasons for doubting that the prebiotic compounds could have been available on the planet.

Biochemistry of Martian Organisms

Life on Earth has two general features: All the processes occurring in living organisms are based on compounds consisting mainly of carbon, and the medium in which all biochemical reactions take place is water. The element carbon is characterized by its ability, either alone or in conjunction with nitrogen and oxygen atoms, to form stable compounds with long chains of atoms, such as proteins, and rings, as in the sugars, pyrimidines, and purines. The suggestion has been made that biological systems based on chains of other atoms, such as nitrogen, silicon, or sulfur, in place of carbon, are conceivable. It is improbable, however, that any such system would have evolved under the conditions existing on Mars.

Furthermore, although it is possible to imagine a biochemistry in which the solvent is liquid ammonia, a hydrocarbon, or even hydrogen fluoride, rather than water, it seems unlikely, in the circumstances, that this would be found on Mars. In 1959, S. L. Miller and H. C. Urey stated: "We know enough about the chemistry of other systems, such as those of silicon, ammonia, and hydrogen fluoride, to realize that no highly complex system of chemical reactions similar to that which we call 'living' would be possible in such media."

It will be assumed that if life has developed on Mars, its chemistry will be somewhat similar to, but not necessarily identical with, that of living systems on Earth. The essential structural and metabolic materials would presumably be carbon and nitrogen compounds and the reaction medium would be water. Consequently, several of the procedures proposed for use in the search for life on Mars, are based on the chemical behavior of life on Earth.

Development of Life on Earth and Mars

Although the average surface temperatures on Mars are appreciably lower than on Earth,

they are probably not so low all over the planet as to have precluded the development of life. Moreover, it is probable that a greenhouse effect, arising from absorption of infrared radiation by gases in the primitive atmosphere, would have served to increase the temperature near the surface by several degrees. On the whole, the temperature would appear not to have been incompatible with the development of life on Mars.

A more serious problem is concerned with the availability of water. It is universally accepted by biologists that life on Earth must have originated in water, such as in the "warm little pond" of Charles Darwin or the "dilute soup" of J. B. S. Haldane. There are several reasons for believing this to be the case. For example, water is by far the most abundant constituent of all living systems. Then, the rough similarity between the ratio of calcium to potassium in sea water and in blood and tissue suggests that life developed, even if it did not originate, in the oceans.

Another argument in favor of the theory that life on Earth developed in water was the need for protection from ultraviolet light from the Sun. The radiation of shorter wavelength, less than about 2400 Å, would probably have been absorbed by the gases of the primordial atmosphere, but ultraviolet rays with wavelengths from 2400 Å up to at least 2900 Å would undoubtedly have reached the surface of Earth. The nucleic acids (DNA and RNA), however, are seriously damaged by radiation in the vicinity of 2600 Å. It is improbable that living organisms utilizing DNA and RNA for replication and for the synthesis of enzymes and other proteins, such as those on Earth do, could have developed unless there was some protection from ultraviolet radiation.

One way in which such protection can be achieved, and probably was on Earth, is to take advantage of the absorption of ultraviolet rays by water. At a depth of several meters in water, the radiation would have been attenuated to such an extent that the nucleic acids would not have decomposed. When oxygen and its associated ozone, both of which absorb ultraviolet radiation, began to accumulate in Earth's atmosphere, the protection offered by water no longer became essential to the continuation of life. Living organisms then emerged from the sea and established themselves extensively on land. It will be seen in due course, however, that other modes of protection from solar ultraviolet rays are conceivable and may well have been, and still be, effective on Mars.

If life on Mars were to have evolved in the same way as on Earth, then a body, or bodies, of liquid water would seem to be essential. Certainly, at the present time, there can be no substantial quantities of liquid water on the planet, but this does not mean that they have never existed. The first conclusions drawn from the Mariner IV pictures of Mars was that the Martian craters were so ancient that there could never have been streams or bodies of water on the planet. Later, it was realized that only the largest craters, which would have taken much longer to erode, were a few billion years old. The smaller craters formed during the early history of Mars have undoubtedly been eroded away completely and have been replaced by others of more recent origin. There is at present no evidence that would disprove the possibility that, at one time, there were substantial quantities of water on Mars.

Because there was nothing in the Martian atmosphere to prevent penetration of solar ultraviolet light of moderate wavelengths, the water would have decomposed in the course of time. The relatively small gravitational force on Mars would have facilitated the escape of hydrogen gas and the oxygen would have oxidized iron compounds on the

319-889 O—69——7

surface to the ferric state. Thus, if liquid water had existed at one time on Mars, its absence now can be accounted for.

There is spectroscopic evidence that the solid material on the surface of Mars contains water of hydration. Some of this water may be present in the form of a hydrated ferric oxide, such as limonite, but there may also be other compounds which contain water in a bound (or chemically combined) form. If this is the case, then there must once have been substantial quantities of water available on the planet. The existence of bodies of water on Mars at some period in its early history cannot be ruled out completely. Such bodies might have lasted for more than a billion years, and this would have been sufficient to permit the evolution of life from prebiological materials.

LIFE IN THE MARTIAN ENVIRONMENT

Persistence of Life Under Martian Conditions

Suppose living organisms had developed on Mars. Could they have continued to evolve under the climatic, atmospheric, and other conditions—which are severe relative to terrestrial standards—that have existed on the planet probably for several hundred million years? Here, again, the answer would be that it is by no means impossible. The main factors that might be expected to affect the persistence of life on Mars are the surface temperature and its variations, the composition and pressure of the atmosphere, exposure to cosmic rays from the Sun and space, the solar ultraviolet radiation, and the lack of water. These factors will be considered in turn. The order of presentation is largely a matter of convenience and does not necessarily represent relative importance.

The estimated surface temperatures on Mars range from a maximum of about 25° C (77° F) at the equator to a minimum of around –100° C (–148° F) or less at the winter pole. Although the temperature variations at any locality would not be quite as large as this, there are, nevertheless, often considerable differences between day and night temperatures, as may be seen from figure 6.22. In some regions at middle latitudes, especially in the summer, the daily temperature ranges from roughly 20° C (68° F) in the daytime to –60° C (–76° F) or so at night. It has been found that certain terrestrial micro-organisms can withstand repeated temperature variations of this kind, although they are not normally exposed to them. A living organism that had evolved on Mars would undoubtedly be even better equipped to survive and reproduce itself under the existing conditions.

The remarkable adaptability of living systems to their environment is shown by the range of temperatures over which they can survive on Earth. Micro-organisms are found near the top of Mount Everest, and others thrive in hot springs where the temperature is near the boiling point of water. What is more, many organisms are known to be capable of existing over a range of temperatures. R. S. Young has reported that bacteria isolated from the hot soil around the volcano Mount Stromboli continued to grow in the cold of a simulated Martian environment. Certain bacteria have been found growing on ice cream and fish stored at temperatures of –10° C (14° F), as well as at room temperature or above. A somewhat similar situation exists for some molds which can still grow in concentrated fruit juices even when cooled to –20° C (–4° F). In laboratory experiments, described below, made to simulate conditions on Mars, some micro-organisms (particularly bacterial spores) survived much larger variations in temperature.

It is of interest that terrestrial organisms,

which normally live at very low temperatures in the polar regions, have evolved in such a manner that the water they contain does not freeze at 0° C (32° F), the normal freezing point. These organisms contain either dissolved salts, as in the halophilic bacteria, or an organic compound, such as a glycerol derivative, found in some insects, which acts as an "antifreeze." These solutes lower the freezing point of water to a temperature which is below that of the surroundings. Consequently, the water in the organism does not freeze even though the temperature falls below 0° C (32° F). Life forms with similar characteristics could have developed on Mars and would be well adapted to an environment in which the temperature fell below the normal freezing point of water every night.

Anaerobic Organisms

The only gases definitely known to be present in the atmosphere of Mars are carbon dioxide, which is a major, possibly the main, component, and small quantities of water vapor. In addition, there may be some carbon monoxide. It would be reasonable to expect some nitrogen and argon, as well as traces of other inert gases that are found in Earth's atmosphere. There is probably a little oxygen gas on Mars, resulting from the decomposition of carbon dioxide by solar ultraviolet radiation, but it will be supposed that the quantity is too small to be of any biological significance.

The first living organisms on Earth must have been anaerobic in nature, because there was probably little or no free oxygen available to them. Such organisms obtained the energy required to carry on their life (metabolic) processes by the oxidation of suitable nutrient material, consisting of carbon compounds, without the use of oxygen gas. There are still a number of types of such anaerobic bacteria known on Earth. Their energy results from the oxidation of carbon (to carbon dioxide) and hydrogen (to water) in organic materials, called substrates; but the oxidizing agent, referred to in general terms as the electron acceptor, is not gaseous oxygen. Some of the known electron acceptors for terrestrial anaerobic organisms and their reduced states, after oxidation of the substrate, are listed below. All of these electron acceptors are probably available on Mars.

Electron acceptor	*Reduced state*
Sulfate (SO_4^{2-})	Sulfide (S^{2-})
Nitrate (NO_3^-)	Nitrogen (N_2)
Nitrogen gas (N_2)	Ammonia (NH_3)
Carbon dioxide (CO_2)	Methane (CH_4)

Of special interest for this planet is that the ferric oxide (Fe_2O_3) in limonite, which is believed to be present on the surface, might well serve as the electron acceptor for oxidizing the organic substrates utilized by living organisms to supply their energy. In 1964, the American biologist, Wolf Vishniac of the University of Rochester, reported that a crude culture of organisms, obtained from river mud, showed vigorous growth in a medium in which freshly precipitated ferric hydroxide (hydrated ferric oxide) was the only electron acceptor. If the organic substrate on Mars is assumed to be a carbohydrate, with the general formula $(CH_2O)_n$, as it would for terrestrial anaerobic organisms, the energy-producing reaction would perhaps be represented by

$$(CH_2O)_n + 2nFe_2O_3 \rightarrow nCO_2 + nH_2O + 4nFeO$$

The carbohydrate is oxidized to carbon dioxide and water, whereas the hydrated ferric oxide is reduced to ferrous oxide.

Terrestrial anaerobic organisms can also produce energy from an organic substrate by the process of fermentation. In a common type of fermentation, the substrate itself (or one of its decomposition products) acts as

the electron acceptor. In other words, the substrate undergoes simultaneous oxidation and reduction. There are usually two main products: one is more highly oxidized than the substrate and the other is in a more reduced state. A simple example is provided by the alcoholic fermentation of carbohydrates; they are first broken down into glucose ($C_6H_{12}O_6$), which is then simultaneously oxidized to carbon dioxide ($2CO_2$) and reduced to ethyl alcohol ($2C_2H_6O$). There would appear to be no reason why metabolic fermentation processes could not occur on Mars.

It has been tacitly assumed that a suitable substrate (or substrates) would be available for Martian organisms. The first heterotrophic living systems on Earth must undoubtedly have utilized material from their surroundings formed during prebiological times. But eventually many autotrophic terrestrial organisms, such as bacteria, algae, and green plants, evolved methods for producing their own substrates from natural materials. The formation of a substrate, such as a carbohydrate, which can be utilized by the organism to yield energy, requires a supply of energy and this is provided by the Sun.

Of particular interest in connection with the possibility of life on Mars are the autotrophic bacteria, which obtain all their nutritional requirements from carbon dioxide and inorganic salts, such as nitrates, phosphates, and others, found in nature. These organisms, however, like green plants, require a supply of energy to permit them to make use of the carbon dioxide in the atmosphere. This energy may be obtained from the Sun in the process of photosynthesis, or it may be derived from chemical reactions.

In photosynthesis, carbon dioxide, the product of carbohydrate oxidation, is reconverted into carbohydrate by means of a reducing material (electron donor), the necessary energy being derived from sunlight. A few different kinds of photosynthesis are known, and in each case the energy is first absorbed from solar radiation by a colored substance. For example, chlorophyll in green plants is utilized in a series of reactions which lead ultimately to the formation of carbohydrates.

The electron donor in green plants and algae is water, and this water is oxidized to oxygen gas in the photosynthetic process. In view of the small (or nonexistent) quantity of oxygen in the Martian atmosphere, it does not seem likely that this type of photosynthesis has developed to any great extent. Alternative electron donors, which play a role in terrestrial photosynthetic bacteria, and the corresponding oxidized states are as follows:

Electron donor	*Oxidized state*
Hydrogen sulfide (H_2S)	Water (H_2O) and sulfur (S)
Sulfur (S)	Sulfate (SO_4^{2-})
Hydrogen gas (H_2)	Water (H_2O)
Organic compound	Oxidized compound

On Mars, there is the possibility that ferrous oxide, formed by the reduction of ferric oxide in the oxidation of a carbohydrate substrate, could act as the electron donor in photosynthesis. The conversion of carbon dioxide into carbohydrate in the presence of light would be represented by the reverse of the equation given above:

$$4n\mathrm{FeO} + n\mathrm{CO_2} + n\mathrm{H_2O} \xrightarrow{\text{light}} 2n\mathrm{Fe_2O_3} + (\mathrm{CH_2O})_n$$

The ferric oxide would be regenerated at the same time. Such a cycle, in which ferric oxide serves as the electron acceptor (oxidizing agent) for the production of energy, whereas the resulting ferrous oxide is the electron donor (reducing agent) in photosynthesis, would be eminently suited to Martian organisms.

Some autotrophic terrestrial bacteria are able to reduce carbon dioxide and store en-

ergy without utilizing sunlight. The required energy in these cases is provided by chemical reaction of the carbon dioxide with a hydrogenous compound. Such chemosynthetic bacteria require oxygen gas, however, for their metabolism. It is not very probable that organisms of this type exist on Mars, but anaerobic chemosynthetic organisms are conceivable. The fact that they do not exist at present on Earth does not eliminate the possibility that they have evolved elsewhere or that they have occurred on Earth in the past.

It may be asked if the low Martian atmospheric pressure, about one-hundredth part of that on Earth, would have any influence on the development of living organisms. There are two aspects to this problem, the first being the matter of the availability of liquid water and the second the pressure itself. At a total atmospheric pressure of 10 millibars, water boils at a temperature of about 7° C (44.6° F). Consequently, liquid water cannot exist for any length of time on Mars except at low temperatures. The significance of this will be considered shortly. As far as the actual low atmospheric pressure is concerned, laboratory experiments with terrestrial micro-organisms indicate that no significant effect is to be expected. Living systems which evolved on Mars would, of course, have become adapted to the low pressure.

Cosmic Rays and Ultraviolet Radiation

Earth is being continuously bombarded by cosmic rays, consisting largely of the nuclei of hydrogen atoms (protons) having high energies. These protons come mainly from space beyond the solar system, but partly from the Sun, especially at times of solar activity as indicated by the formation of flares. In sufficient quantity, the high-energy protons and other particles of cosmic rays can have harmful effects on living systems. Fortunately, the terrestrial magnetic field and the atmosphere serve to reduce the cosmic-ray intensity to harmless proportions at Earth's surface. In terms of the conventional unit, the rad,[1] which is used to express the dose, or dose rate, of such nuclear radiations, the average total annual dose at sea level caused by cosmic rays is about 0.035 rad. At an altitude of 5000 feet, it is roughly 0.050 rad. These values may be compared with about 0.09 rad per year from radioactive elements in the ground.

Because the magnetic field of Mars is either very weak or nonexistent, and the atmospheric density is low, the intensity of cosmic rays on the surface of this planet is much greater than on Earth. According to estimates made in 1965 by J. J. O'Gallagher and J. A. Simpson, the radiation dose from cosmic rays at the Martian surface is equivalent to more than 7 rads per annum. Although this is about 200 times as large as the average dose on Earth, it is not likely to be a serious hazard to possible micro-organisms on Mars.

A radiation dose of something like 600 rads received by a human being over the whole body within a period of a day would prove fatal in most cases. If the same dose were spread over a longer period of time, the biological effects would be much less severe. It may be significant from the evolutionary standpoint that the simplest organisms—bacteria, protozoa, algae—are much less sensitive to nuclear radiation than man and vertebrates in general. It requires a dose of 300 000 rads or more, received in a short time, to kill the simple organism paramecium. Even for the relatively radiosensitive bacteria, such as *Escherichia coli,* found in the colon, the lethal dose is 5000 rads. A cosmic-ray dose of about 7 rads per year (0.02 rad per day) would thus

[1] One rad is equivalent to the absorption of 100 ergs of energy per gram of absorber.

have no apparent effect on terrestrial micro-organisms. Life forms which evolved on Mars would probably be even less sensitive than those on Earth.

It has been pointed out by C. Sagan, however, that although a dose rate of 7 rads per year would not be immediately fatal to micro-organisms, and even to some higher forms of life, such a radiation exposure would result in the production of a large number of mutations (p. 159). Because most mutations, at least on Earth, are known to be deleterious, they might have led to the extinction of life on Mars even if it had been present at one time.

A much more serious hazard than cosmic rays on Mars, at least to life of the terrestrial type, would be ultraviolet rays from the Sun. The blue haze described in chapter VII may possibly be effective in reducing the light of short wavelength reaching the surface. But this is far from being certain. It must be assumed, therefore, that essentially all the ultraviolet radiation coming from the Sun, at least in the wavelength range of 2400 to 2900 Å, reaches the surface. The best estimate is that ultraviolet energy falls on Mars at the rate of 2000 erg/sec/cm^2 (2×10^{-4} watt/cm^2) of surface.

Tests made with terrestrial organisms show that most are killed in a short time when exposed to the ultraviolet radiation intensity expected on Mars. The spores of some fungi, however, especially those having a red or black pigment, appear to be exceptionally resistant to ultraviolet rays, and have survived exposure to large doses of this radiation. Some other micro-organisms are able to survive with a minimum of protection, such as may be provided by small particles of soil or even rough surfaces. Incidentally, soil is a very good absorber of ultraviolet radiation and, if they did not have to depend on photosynthesis to provide energy and nutriment, organisms on Mars could live a short distance below the surface and be completely protected.

Another possibility is that Martian organisms have developed natural defenses against ultraviolet light. In the report, "Biology and the Exploration of Mars," prepared under the auspices of the U.S. National Academy of Sciences-National Research Council and published in 1965, three types of possible defense mechanisms are considered. First, "highly absorbent [i.e., capable of absorbing ultraviolet rays] organic material may be incorporated in the cell wall or in a waxy cuticle. . . . The radiation so absorbed [is converted into heat and] would . . . serve to raise the temperature of the organism."

A second concept is that "an absorbent inorganic material, such as the limonite of Martian soil, could be combined with . . . [a] silica shell to make a rigid iron glass with strong ultraviolet absorbing properties." It is worth noting in this connection that the window glass in common use, which usually contains sodium and calcium silicates, is a good absorber of ultraviolet radiation. It is thus possible to conceive of an iron-silicate glass (or similar material) which would remove ultraviolet radiation but would permit the passage of light of longer wavelength for photosynthesis.

The third suggestion is that "an organism [on Mars] might be shielded by fluorescent material and the emitted light could support photosynthesis." In explanation of this possibility, it should be mentioned that a fluorescent substance is one that absorbs radiation of relatively short wavelength, for example, in the ultraviolet region, and subsequently emits part of the energy as radiation of longer wavelength, for example, in the visible portion of the spectrum.

It is indicated in the report mentioned above that simple amebalike organisms belonging to the *Testacidae,* such as *Arcella* and

Difflugia, which have substantial tests (shells), could serve as models for shielded Martian organisms. These ideas are speculative, but they do indicate possibilities which are by no means unreasonable.

The Importance of Water

The most serious factor capable of affecting life on Mars is the present shortage of water. Because of the low pressure of water vapor in the atmosphere, bodies of liquid water could not exist for an appreciable time even when the surface temperature is above the normal freezing point of 0° C (32° F). Solidified water, in the form of hoarfrost, would tend to vaporize rather than melt to liquid when the ground and atmosphere of Mars warm up in the spring. There may be some localities, however, where liquid water could have a transient, nonequilibrium existence. The only satisfactory explanation, although not necessarily the correct one, proposed to account for the dark-colored band which appears around each retreating polar cap on Mars is that it represents ground wetted by liquid water.

It is known that in the absence of water, spores of terrestrial bacteria can remain dormant for long periods. When water is available, the spores will germinate. Something similar may be happening on Mars. Furthermore, lichens, which grow on bare rocks, and some bacteria and molds, which survive in dry flour, obtain their water requirements from vapor in the air. There is more water vapor even in the driest atmosphere on Earth than in the Martian atmosphere, but the evolution of organisms on Mars could have fitted them to the prevailing conditions in this respect as well as in others. Many terrestrial species, for example, have become adapted to living in dry environments. A well-known instance is the kangaroo rat which can survive for long periods without liquid water. It obtains all the water it needs by the oxidation of the carbohydrates in its food.

The situation may be summarized once more in the words of N. H. Horowitz:

> It is certainly true that no terrestrial species could survive under average Martian conditions . . . except in the dormant state. But if we admit the possibility that Mars once had a more favorable climate which was gradually transformed to the severe one . . . [of] to-day, and if we accept the possibility that life arose on the planet during this earlier epoch, then we cannot exclude the possibility that Martian life succeeded in adapting itself to the changing conditions and remains there still.

There is another aspect, although a speculative one, concerning water on Mars that may have a bearing on the occurrence of living organisms on the planet. Water might be trapped below the surface in the form of permafrost (p. 83). J. Lederberg and C. Sagan suggested in 1962 that there may be local regions on Mars where heat from the interior reaches the surface. The permafrost layer in these regions would melt and form liquid water. As a result, there would be moist soil, at least, or possibly a pond or a small lake, equivalent to a hot spring, which is continuously fed by melting permafrost. Living organisms could then thrive in such locations. Volcanic action on Mars is thought to be less than on Earth, because of Mars' smaller size. It is not at all unlikely that there is such activity at a sufficiently moderate depth to cause heating of the surface in some places.

Experimental evidence for the formation of permafrost has been obtained by R. S. Young and his collaborators by means of a Martian simulator constructed at the Ames Research Center of the National Aeronautics and Space Administration. The simulator consists of a closed, heavily insulated, aluminum box containing a mixture of equal parts of limonite and sand to represent the Martian soil. The temperture is controlled by passing liquid nitrogen at the required rate through a

heat exchanger at one end, representing the north pole, and by means of infrared lamps at the other end, representing the equator. An electric-clock mechanism automatically turns the lamps on and off in such a way as to simulate the day and night temperature variations. Martian seasons can be imitated by raising and lowering the infrared lamps.

"During operation, the simulator was brought to equilibrium over a period of several weeks, during which the nighttime pole temperature is about $-100°$ C and the equatorial temperature varies from a high of $+30°$ C at noon to $-70°$ C at night. These extremes are considered to be most likely for a Martian equatorial summer day." Moisture present in the simulated atmosphere and in the soil condensed as frost at the cold (polar) end of the box. Dry gas was then cycled above the soil in the chamber until a light frost, comparable in appearance to what is seen at the Martian pole, remained at the cold end.

The box was then opened and soil removed to determine whether or not ice was present below the surface. It was found that under these conditions an ice layer, equivalent to permafrost, is formed. It is fairly deep at the equator but comes to the surface at the pole (fig. 10.1). It is of interest that the subsurface temperature in the simulator was constant at $-50°$ C; that is, 223° K. This is not very different from the value of approximately 210° K estimated from microwave measurements (p. 134).

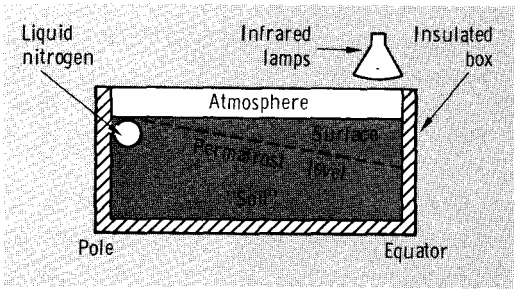


FIGURE 10.1. Martian simulator and formation of permafrost.

If there is a permafrost layer beneath the Martian surface and if there are warm regions where liquid water is formed, evaporation would take place continuously into the dry atmosphere at a rapid rate, especially in the daytime. It is possible, therefore, that the water might become quite concentrated in various mineral salts. But this should prove no detriment to properly adapted life forms. Certain terrestrial, halophilic, bacteria, for example, live and multiply in salt water and even in saturated salt solutions such as exist in the Dead Sea. These organisms are also able to tolerate bromides which are toxic to most other life forms. Some organisms are known to live in highly acid solutions, whereas others are normally found in fairly alkaline media. The adaptability of the same fundamental biological processes to a very wide variety of environments is quite remarkable.

Terrestrial Organisms in the Martian Environment

At the suggestion of Hubertus Strughold of the U.S. Air Force Aerospace Medical Division, laboratory tests were initiated around 1957 to determine if terrestrial organisms could survive and grow under simulated Martian conditions. These studies had two objectives: to determine if organisms carried from Earth on a spacecraft would continue to live on Mars and thus contaminate the planet, and to obtain information concerning the possibility of indigenous life on Mars. Several investigators found that a variety of anaerobic terrestrial bacteria and spores would be expected to survive on Mars, and that they would even multiply in a nutrient medium if sufficient moisture were available.

Because of the uncertainties concerning

the composition and pressure of the Martian atmosphere, most of the work done before 1965 had to be repeated. The general conclusion, however, remains unchanged. It is that terrestrial spores and micro-organisms could survive on Mars, and that the limiting factor for growth and multiplication is the amount of moisture in the soil. A highly interesting observation was reported in this connection in 1967 by E. J. Hawrylewicz and his associates in the United States. They found that with large daily (day-night) temperature changes, from 25° to −65° C (77° to −85° F), accompanied by alternate thawing and freezing of the moisture in the soil, some micro-organisms would grow at soil moisture concentrations for which growth would cease when the temperature remained constant. The large diurnal temperature variations which occur on Mars might thus prove advantageous where the amount of water is limited.

It should not be construed from the information presented above that organisms which might exist on Mars would necessarily be similar to the simplest organisms on Earth. Because the only terrestrial anaerobic species are micro-organisms, it is inevitable that laboratory experiments under simulated Martian conditions must be made with such bacteria. The fact which has been established is that, provided some water is available, there is no reason why life on Mars should not be possible.

Because of the greater intensity of cosmic rays and ultraviolet light on Mars, the frequency of mutations, which both of these types of radiation are known to produce, would be significantly greater than on Earth. Taking into consideration the different environmental conditions, life forms could have evolved on Mars that are inconceivable from the terrestrial standpoint. In view of the absence of oxygen gas in the atmosphere, the high intensity of ultraviolet radiation, and the shortage of water, it seems probable that living organisms on Mars could be quite different from those which now exist on Earth.

Possible Evidence of Life on Mars

This discussion leads to the conclusion that life on Mars is not impossible. The question may now be asked: Is there any evidence that there are living organisms on Mars? Some scientists would say there is, but others would say there is none. Apart from the somewhat fanciful and improbable idea that the "canals" on Mars are the work of intelligent beings, there are three main lines of evidence that have been suggested as indicating the possible existence of one or more forms of life on the planet.

First, there is the wave of darkening which spreads from the polar regions to the equator during the local spring and summer in each hemisphere of Mars. Until recent years, the only rational explanation of this phenomenon appeared to be that vegetation of some kind which had been dormant during the winter revived under the influence of higher surface temperatures and, especially, the availability of water in some form. Serious consideration must now be given, however, to the alternative interpretation presented in chapter VI, based on the transfer of dust particles by the prevailing winds.

One of the strongest arguments for life on Mars was at one time thought to be the recovery of the dark regions after the dust storms that envelop large areas of the planet from time to time. If, as has been suggested, the dark portions of Mars are elevated regions, rather than depressed ones as had long been assumed, then the phenomenon could well be accounted for by the scouring effect of winds.

A third hypothesis is based on the spectro-

scopic observations made in 1956 and 1958 by W. M. Sinton. He found that the infrared spectra of dark areas, such as Syrtis Major, exhibited distinct absorption bands in the regions of 3.43, 3.54, and 3.69 μ wavelength (p. 76). From the bright areas, however, the absorption appeared to be much less distinct. It is well known that organic molecules containing C—H bonds give characteristic spectra in the wavelength region of 3.4 to 3.5 μ. Consequently, the absorption bands observed by Sinton were attributed to the presence of organic compounds produced by life forms existing on the dark areas of Mars.

The Sinton bands were the subject of much controversy until an alternative explanation of their origin, in which Sinton himself concurred, was proposed by D. G. Rea and B. T. O'Leary in 1965. Two of the three bands, those at 3.58 and 3.69 μ, are caused by HDO molecules in the terrestial atmosphere, HDO representing a molecule of water (H_2O) in which one atom of hydrogen (H) is replaced by an atom of the less common isotope deuterium (D). These two absorption bands are spurious and have nothing to do with Mars. There remains, however, the third band at 3.43 μ, which is really more typical than the others of the C—H linkage. This band is very weak and its significance is uncertain. It may be due to organic compounds or even to carbonates on the Martian surface.

It has sometimes been stated that the photographs of Mars taken by the Mariner IV spacecraft do not indicate the presence of life. This statement is quite correct as far as it goes. Actually, the Mariner IV pictures do not provide, and were not expected to provide, any evidence regarding the existence of living organisms on the planet. The distance of closest approach of the spacecraft to the Martian surface was 9846 kilometers (6118 miles), and photographs of Earth taken at this distance would show no signs of life. In fact, apart from occasional glimpses of human activity, such as the wake of a ship at sea, a newly completed broad highway, and a grid-like pattern of trees remaining after a logging operation, there are no indications of life in the pictures of Earth taken from Tiros and Nimbus satellites which sometimes approached as close as 500 kilometers (300 miles) or so to the surface.

THE COLONIZATION OF MARS

Introduction

Although it will probably not be in the foreseeable future, the situation may eventually arise when conditions on Earth will be such that colonization of the planets becomes desirable or even necessary. Possible factors in this regard are overpopulation, extensive radioactive (or other) contamination, flooding of large areas by the melting of Earth's icecaps, or even, at the other extreme, a new ice age. The obvious, and perhaps only, other planet within the solar system that might be considered for human habitation is Mars.

Mars is the second closest planet to Earth and on favorable occasions it is not much farther away (56 million kilometers) than the closest, Venus (40 million kilometers). The days and nights on Mars are approximately the same length as on Earth, and there are similar seasons, spring, summer, autumn, and winter, although they are much longer in duration on Mars. The daytime surface temperatures in a belt about 2500 kilometers (1500 miles) wide around the equator are not greatly diferent from those on many parts of Earth, but the Maritan nights are everywhere extremely cold by terrestrial standards.

By the time human beings are ready to leave Earth in large numbers, the problems associated with travel over considerable dis-

tances in space will undoubtedly have been solved. Transportation to Mars will, therefore, not present any serious difficulties. And living on the planet, at least in the early stages of its colonization, will not be very much different from living in a large space vehicle for extended periods, except that the structure will be at rest on the ground.

In particular, the lack of oxygen and the low atmospheric pressure on Mars will make it necessary for man's habitations to be closed structures. Such structures can, however, be quite large, so that there is ample room for movement. They must not be thought of in terms of the cramped quarters of the Apollo spacecraft scheduled to make the earliest manned expeditions to the Moon. In fact, apart from the requirement for the recovery and reutilization of wastes, the conditions in an enclosed building on Mars would be comparable with those in a large jet aircraft flying at a high altitude above Earth. Incidentally, although the environment is entirely different, knowledge gained from living in closed vessels (submersibles) under water will be very useful in establishing a colony on Mars.

In order to venture outside the closed structures on Mars, men would have to wear pressure suits such as are now planned in connection with the exploration of the lunar surface. These suits will include a supply of oxygen and a means for removing exhaled carbon dioxide and excess moisture. The extensive experience that will have been obtained in many lunar and planetary landings, with increasingly larger groups of astronauts, will render the use of pressure suits relatively commonplace.

Requirements for Human Habitation

In his report entitled "Habitable Planets for Man," published in 1964, Stephen Dole of the Rand Corp. has stated: "Conditions of temperature, light, gravity, atmospheric composition and pressure, and water are probably the major human requirements." Apart from excessive ultraviolet radiation and the cold nights, the temperature, light, and gravity on Mars would probably be acceptable to human beings from Earth. The atmospheric composition and pressure and the need for water, however, would present special, but not necessarily insuperable, problems.

The five major requirements given above will be examined with reference to the conditions on Mars. There are actually two different aspects of these requirements: as they affect man himself, and their relation to the production of food. Both of these aspects will be taken into consideration in the following discussion.

As already mentioned, the temperatures during the day on Mars, in areas not far from the equator, would be tolerable to human beings, animals, and plants. But the nighttime temperatures are very low. This would simply mean that the closed structures would have to be heated at night. Reference will be made later to possible sources of power, but if sufficient energy is available, maintaining a comfortable temperature at night would not appear to be very difficult. It is assumed that the structures would be well insulated.

Adequate light is required for human beings to see by and for the photosynthetic process in green plants in the production of food. The total amount of solar radiation falling on Mars per unit area in a given time ranges from 36 percent (at aphelion) to 52 percent (at perihelion) of that reaching Earth. The density of the Martian atmosphere is much less than the terrestrial atmosphere, and there are probably, on the whole, fewer clouds. The average light intensity is not greatly different on the two planets and the sunlight on Mars should thus be sufficient for both plants and man.

The essential difference between the radiations from the Sun falling on the surface of Mars and Earth is that there may well be more ultraviolet in the former case. Ultraviolet light is harmful to all terrestrial organisms. But because enclosed structures or pressure suits would be mandatory, in any event, protection against ultraviolet radiation would be a simple matter. All common materials that are opaque to visible light will absorb ultraviolet rays. Furthermore, as already noted, common window glass is largely opaque to ultraviolet even though it is transparent to light in the visible range.

The proper growth of plants on Earth requires a more-or-less definite pattern of alternating light and dark intervals, as provided by day and night. Because the period of rotation of Mars is approximately the same as for Earth, the lengths of days and nights, at equivalent latitudes, are similar. Plants also require seasonal variations in the length of the day, as well as in the temperature. On Mars, these variations will not be the same as on Earth. It may be possible to develop plants which are adapted to the conditions on Mars, but if this is not practicable the creation of artificial seasonal variations in daylight and temperature, to correspond to those on Earth, appears feasible.

For man's convenience of movement, on the one hand, and the avoidance of fatigue, on the other hand, the force of gravity should be neither too small nor too large compared with the terrestrial gravitational force. It seems that the conditions on Mars, with a gravitational acceleration about 38 percent of that on Earth, should be quite tolerable. Tests made in connection with the lunar landing program have shown that man can learn to move about and perform various tasks in a gravitational field that is less than half as strong as on Mars. Relatively little is known about the effect of low gravity on plant growth, but the indications are that it is not detrimental.

The atmosphere on Mars, with regard to both composition and pressure, is quite unsuitable for human beings and animals. Consequently, an artificial atmosphere will have to be provided, as indicated earlier. It is possible that some of the oxygen required for breathing can be extracted from the limonite (or other form of ferric oxide) which is believed to be present on the Martian surface. A better alternative, however, would be to take advantage of the abundant supply of carbon dioxide in the atmosphere. As a result of photosynthesis by green plants, this could be converted into oxygen which would be used by man. Admittedly Mars provides few positive features for human habitation, but the atmospheric carbon dioxide is one of them.

Finally, there is the problem of an adequate supply of water. This may well prove to be the limiting factor as far as colonization of Mars by man is concerned. Of course, if there is indeed a layer of permafrost at a moderate depth, as has been surmised, a supply of water would be available. Minerals on or near the surface evidently contain some water of hydration, but its extraction would require a considerable expenditure of energy. Because many materials needed on Mars would have to be brought from Earth, water might be included. Another possibility would be to bring hydrogen and then produce water—and energy—by chemical reaction with the ferric oxide on the Martian surface. In any event, careful conservation of water would have to be practiced on Mars.

Food and Power Supplies

Two requirements of human life, which have been implied in the foregoing, are food and a source of power. There may be indige-

nous material on Mars that is suitable for food, but it must be assumed that such will not be the case. If green plants can be grown, as is quite conceivable, assuming mineral elements, nitrogen, and water are available, it should be possible to raise animals for food purposes. Such animals could not be grazed in large open spaces. Production would have to be carried out in the factorylike manner now used on Earth for chickens, eggs, and milk. Incidentally, it is not at all impossible that, in the not too distant future, nutritive materials, which have both the taste and protein content of meat, will be made from primitive plantlife, such as algae. Animals will then be unnecessary.

There are two natural sources of power on Mars which might possibly be utilized: solar radiation and wind. Although radiation from the Sun would be reliable, the Martian winds appear to be uncertain and sometimes attain high velocities. Combustible fuels, such as gas, oil, and coal, would be useless because of the lack of oxygen. Nuclear energy, either by fission or fusion, would seem to be the best solution to the power problem, although the fuel would probably have to be brought from Earth. Ultimately, uranium, to serve as a fission fuel, might be extracted from minerals on Mars. There is little doubt that such minerals are present on the planet.

Conclusion

The general conclusion to be drawn, at present, is that Mars is not an attractive, but not an impossible, planet for human habitation. The colonization of Mars would probably be undertaken only under conditions of dire emergency. Nevertheless, such conditions might arise in the future. Scientists should consider, with the knowledge available to them, the steps which would have to be taken to permit a considerable population to live on Mars. In this respect, knowing more about Mars itself is, of course, of paramount importance.

FOR FURTHER INFORMATION

DOLE, S.: Habitable Planets for Man. The Rand Corp., R–414–PR, 1964.

HAWRYLEWICZ, E. J.; ET AL.: Probability of Growth of Viable Micro-organisms in Martian Environments. Life Sci. Space Res., North-Holland Pub. Co., vol. VI, 1968.

HOROWITZ, N. H.: The Search for Extraterrestrial Life. Science, vol. 151, 1966, p. 789.

KILSTON, S. O.; DRUMMOND, R. R.; AND SAGAN, C.: A Search for Life on Earth at Kilometer Resolution. Icarus, vol. 5, 1966, p. 79.

LEDERBERG, J.; AND SAGAN, C.: Microenvironments for Life on Mars. Proc. Natl. Acad. Sci., vol. 48, 1962, p. 1473.

MAMIKUNIAN, G.; AND BRIGGS, M. H., EDS.: Current Aspects of Exobiology. Pergamon Press, 1965.

MICHAUX, C. M.: Handbook of the Physical Properties of the Planet Mars. Ch. 18, NASA SP–3030, 1967.

PACKER, E.; SCHER, S.; AND SAGAN, C.: Biological Contamination of Mars II: Cold and Aridity as Constraints in the Survival of Terrestrial Microorganisms in a Simulated Martian Environment. Icarus, vol. 2, 1963, p. 293.

PITTENDRIGH, C. S.; VISHNIAC, W.; AND PEARMAN, J. P. T., EDS.: Biology and the Exploration of Mars. Natl. Acad. Sci.-Natl. Res. Council Publ. 1296, 1966.

SALISBURY, F. B.: Martian Biology. Science, vol. 136, 1962, p. 17.

SALISBURY, F. B.: The Possibilities of Life on Mars. Proc. Conf. Explor. Mars. Virginia Polytechnic Institute Engineering Extension Series, Circular no. 5, sec. V, 1965, p. 15.

SHKLOVSKII, I. S.; AND SAGAN, C.: Intelligent Life in the Universe. Holden-Day, 1966.

SHNEOUR, E. A.; AND OTTESEN, E. A., COMPS.: Extraterrestrial Life: An Anthology and Bibliography. Natl. Acad. Sci.-Natl. Res. Council Publ. 1296A, 1966.

YOUNG, R. S.: Extraterrestrial Biology. Holt, Rinehart & Winston, 1966.

XI

The Detection of Life on Mars

BIOLOGICAL EXPLORATION OF MARS

Introduction

THE SEARCH FOR LIFE on Mars actually involves much more than trying to determine if any living organisms are now present on the planet. In a report entitled "Space Research: Directions for the Future," issued by the Space Science Board, National Academy of Sciences-National Research Council in 1966, the important point is made that—

> in planning the biological exploration of other planets, the question "Is life present?" should be broadened to the question, "Is any stage of biological evolution observable?" This question covers the search for evidence of early stages in the evolution towards organisms (the prebiotic stage) as well as for living organisms and for the remains of extinct life.

As has been pointed out by C. Sagan, "there is a continuum of possibilities from no life (or life-related chemicals), to prebiological organic matter, up through complex life forms."

The one certain way of answering the question posed above concerning the stages of biological evolution on Mars would be to land man or devices on the surface of the planet to collect samples of the soil in various locations and return them to Earth for detailed study. It is quite possible that the situation will not be resolved completely until this difficult task is achieved. In the meantime, however, attempts will be made to obtain relevant information by means of suitable instruments on spacecraft.

Three general types of space vehicles will be used in the exploration of Mars. These are, in order of increasing complexity: flyby devices, such as the Mariner spacecraft; orbiters; and landers (ch. XIII). It is doubtful that any information concerning Martian biology can be obtained from flyby vehicles. On the other hand, the most positive results are to be expected from landers. Nevertheless, some use can perhaps be made of an intermediate phase involving spacecraft which orbit the planet but do not land on it.

In 1966, a Working Group on Mars Orbiters, which made a study for the Biosciences Program Division of the National Aeronautics and Space Adminstration, reported that—

> the direct detection of life on Mars from an

> orbiter vehicle is considered unlikely, but there are promising possibilities for determining physical parameters of biological relevance as boundary conditions on the ecology of Martian organisms, and for detecting phenomena that may be due to the activities of . . . [these] organisms.

Some of the appropriate studies that can be made with instruments carried by Mars orbiters are described below. Most of this chapter, however, will be devoted to the search for life, or related information, by means of spacecraft landed on the surface of the planet.

Use of Orbiters

The main use of orbiting vehicles in connection with the biology of Mars would be to study the characteristics of various parts of the surface and the seasonal changes they exhibit. By correlating the results with simultaneous observations of the atmospheric composition, it may be possible to draw conclusions of interest. To be of significant value in this respect, however, the orbiter should have a lifetime of 120 to 180 days, at least. This requirement places limits on the altitude and nature of the orbit, and on the payload of scientific instruments that can be carried by the spacecraft.

The instruments may include a television camera which could view the dark areas of Mars and observe the changes in appearance that occur in the local spring and summer. As seen in earlier chapters, such changes may, or may not, be related to life on Mars. The amount of detail visible in the television pictures would depend on the facilities for transmitting the information from the orbiter and for receiving it on Earth.

The use of the polarization of scattered sunlight to indicate the nature of the Martian surface material was described in chapter VI. In this connection, the variation of the polarization with the phase angle is an important factor in the identification (fig. 6.6). When polarization measurements of the surface of Mars are made from Earth, however, the maximum phase angle, around quadrature, is only 47 degrees (p. 44), and is often less, depending on the positions of Earth and Mars in their respective orbits. Better discrimination among different possible surface components could be made if the polarization was measured over a larger range of phase angles.

By means of an orbiting spacecraft, the phase angle could be increased, in principle, to 180 degrees; that is, 90 degrees each side of the line joining Mars and the Sun. Much more precise identification could thus be made of the Martian surface material than is possible from Earth. In general, such polarization measurements are expected to have the capability of suggesting the composition of the substances covering the bright and dark regions of the planet, and of providing information about the granularity of the dark areas. The latter feature may well be related to the presence of living organisms on Mars.

An infrared radiometer on board an orbiter could indicate the surface temperatures at many locations at different times of day and at different seasons of the year. An infrared spectrometer would provide information concerning the surface material and the gases in the adjacent atmosphere.[1] Many organic compounds found in living organisms have characteristic infrared spectral features by which they might be identified. In addition, the presence of the gases methane and nitrous oxide in the Martian atmosphere would be significant, because these gases in the terrestrial atmosphere are of biological origin.

The infrared radiometer could measure the temperature of the surface of Mars during the local night, which is something that cannot be done from Earth. At present, the

[1] The infrared radiometer and spectrometer are described more fully in ch. XIII.

nighttime temperatures, such as those in figure 6.22, are inferred from the observations made of those parts of the Martian surface that are exposed to the Sun. The temperatures at night determine the extent of condensation of water from the atmosphere, and so they may have a bearing on the possible existence of life. Furthermore, the infrared radiometer might detect "hot spots" where the conditions are expected to be favorable for the existence of living organisms.

It was seen in chapter VI that, whereas an infrared radiometer indicates the surface temperature, a microwave radiometer gives the temperatures below the surface. Such measurements might provide a better identification of hot spots than would surface temperatures. The polarization of the microwaves could provide an indication of the surface roughness and its seasonal variations. Like the granularity, this might be of biological significance.

Much information concerning the physical characteristics, and perhaps the nature, of the Martian surface could be obtained from a study of radar reflections. It appears that the potential returns would not justify the inclusion of a relatively massive radar transmitter and receiver on an orbiter. A proposed solution to the problem is to use a powerful radar transmitter on Earth and to study the reflections by means of a relatively lightweight receiver on an orbiting spacecraft.

Use of Landers

The first direct tests for life, past or present, as well as for prebiological compounds, on Mars will be made with instruments that will be landed on the surface of the planet. Any instrument to be used in this connection must satisfy a number of basic requirements. For reasons which will be explained in chapter XII, the instruments must remain effective even after a drastic sterilization procedure designed to kill all terrestrial micro-organisms including spores. The instruments must also be compact and light in weight so that several can be carried by a single vehicle, and they must be capable of operating automatically or by remote control from Earth. In addition, they must be rugged enough to remain undamaged after being dropped to the surface from a spacecraft, and they must not be affected by the large daily temperature changes. Furthermore, information obtained from the instruments must be in a form that can be relayed to Earth by radio telemetry.

An important feature of instruments designed to test for life on Mars is the capability of taking satisfactory samples of material for examination. A number of ideas have been suggested, including the use of sticky string to which surface particles will adhere, various types of suction devices, and the formation of suspensions of very small particles in a gas (aerosol). A preference for small particles is indicated by the fact that, on Earth, the number of micro-organisms found in a given weight of soil is greater the smaller the particle size. Most of the proposed schemes for obtaining samples, however, would take material at or close to the surface. It is not at all improbable, however, that if living organisms exist in the Martian soil, they will be found below the surface where they are protected from ultraviolet radiation and where the amount of water may exceed that nearer to the surface.

Principles of Life Detection

Apart from the foregoing essentially mechanical conditions, the instruments must be designed to provide specific information that is relevant to the problem of interest. From a detailed analysis made in 1965, R. S. Young and his collaborators concluded that

"for the purposes of life detection, chemistry, metabolism, and reproduction seem to be the best attributes of life upon which to build an experimental program." To these three aspects of living organisms, one more, morphology, may be added because of the general, rather than specific, information it can provide.

A study of the chemistry of the Martian surface would involve a search for complex organic compounds of biological interest, such as nucleic acids and proteins, or of their simple components, including pyrimidines, purines, amino acids, and sugars. Even if Martian organisms are based on entirely different compounds than is terrestrial life, it is probable that they would nevertheless be polymeric molecules of high molecular weight; that is, macromolecules. Determination of molecular weight and also of the types of atomic linkages or groupings present would thus be useful. Optical activity is another chemical characteristic of the sugars and amino acids, with the exception of glycine, found in terrestrial organisms. The discovery of substances on the Martian surface capable of rotating the plane of polarization of light would be highly indicative of life, past or present.

Metabolism refers to the processes by which the organism obtains and utilizes the materials and energy required to maintain and reproduce itself. A decrease in concentration of a food material (substrate) or an increase in a waste product, carbon dioxide, for example, would be signs of metabolism. An increase in temperature of a closed system might possibly be used to indicate that metabolic processes are taking place. Furthermore, such processes in terrestrial organisms are usually accompanied by a change in the acidity or alkalinity; that is, in the hydrogen ion concentration (or pH) of the medium. Thus, many of the changes associated with metabolism can be observed and measured.

"The demonstration of reproduction," say Young and his associates, "would certainly be a dramatic means of life detection, and indeed, if it could be unequivocably shown, would probably be the most convincing." Unfortunately, the detection of reproduction might prove to be very difficult. The most unambiguous means for observing the process would be a television microscope, because it is very improbable that reproduction of anything but a micro-organism could be detected. If small particles of similar size and form were actually seen to increase in number, there would be little doubt that a living system was under observation. A simpler, but less decisive, proof of reproduction would be the formation of turbidity in a previously clear nutrient medium.

The term morphology is used to refer both to the external form and internal structure of organisms. If life forms on Mars are similar in appearance to those on Earth, they could easily be recognized by microscopic or other visual examination. Although this situation will probably not arise, there are related ways in which living organisms might be detected. Apart from the overall form, there may be indications of an internal structure quite different from that of inorganic materials. The latter on Mars will, of course, be the same as on Earth, and so their form and structure will be well known. Perhaps the most convincing morphological evidence of life would be observation of spontaneous changes such as motion as a whole or of constituent parts, or changes resulting from alterations in the surrounding medium.

In devising procedures to be used in the search for life on Mars, scientists are venturing into the unknown. One problem, for example, is the type of organism to look for. Most of the tests for life are based on the as-

sumption that, if there are forms of life on Mars, they will include micro-organisms. In the words of the eminent biologist Joshua Lederberg, "we can imagine another world with only microbes, but we cannot conceive of one lacking microbes if it bears any life at all." There may, of course, be more advanced life forms on Mars, but if they exist they would probably be identified later.

The only kind of life known to scientists at present is, of course, life on Earth. Hence, it is obvious that the techniques proposed to seek evidence for life on Mars are such that terrestrial organisms would respond in a positive manner to all the tests applied. Whether or not life that had evolved on Mars would react in the same way is, however, quite uncertain. Because one or two tests may yield misleading results, the reasonable approach is to devise as many different tests as possible.

The scientific principles underlying a number of suggested experimental procedures, which appear to be promising for operation on Mars, are outlined in the following sections. Although some specific instruments have been designed for testing on Earth, it is probable that they will undergo considerable change before there is an opportunity to land them on the Martian surface. Consequently, little will be said about this aspect of the problem. In any event, it is doubtful that it will be possible to make use of all the proposed instruments. The decision as to which will and which will not be landed on Mars, and the form they will take, cannot be made until much more work has been accomplished.

CHEMICAL METHODS OF LIFE DETECTION

Gas Chromatograph

The gas chromatograph is basically an instrument for separating the gases or vapors in a mixture. The individual gases present are identified and their relative amounts determined by comparison with samples of known substances. The use of gas chromatography for detecting life (or life-related compounds) on Mars was proposed in 1963 by V. I. Oyama of the National Aeronautics and Space Administration's Ames Research Center, and a system suitable for landing on the surface of the planet is being developed in the United States by the Jet Propulsion Laboratory in conjunction with S. R. Lipsky and with J. E. Lovelock.

The apparatus used in gas chromatography consists essentially of a column, 10 or more feet in length, twisted into a helical coil for compactness. It is filled with a powdered porous solid, such as a silica powder, and maintained at constant temperature. In gas-solid chromatography the powder itself serves to hold, or absorb, the gas to be analyzed, whereas in gas-liquid chromatography, which is the preferred technique, a nonvolatile liquid, such as a silicone oil, is included in the column to act as a solvent for the gas.

In both procedures, a sample of the gas mixture to be analyzed is forced through the column by means of an inert carrier gas, usually helium or argon. The gases in the mixture are then separated in accordance with the strength with which they are held, either by adsorption or in solution, in the column. The constituent that is least strongly held passes out of the column first and the one most strongly held emerges last, with the others in order in between.

In practice, the carrier gas is allowed to flow through the column at a constant controlled rate until a steady condition is established. The sample to be analyzed is then injected into the stream of carrier gas at the head of the column. The times at which the various constituent gases of the mixture leave the column, that is, the retention times, and

their respective amounts are recorded by means of a detector. A schematic representation of the operation of a gas chromatograph is given in figure 11.1.

A number of different types of detectors have been employed. Two of special interest are those which measure the thermal (or heat) conductivity and those which indicate the ionization properties of the gas. The detector actually records the difference in the particular property of the carrier gas alone and the carrier gas plus the component extracted from the sample. As a result, it indicates the magnitude of the property, and hence the amount, of the extracted component passing through it at a particular time.

For use on the surface of Mars, a simple and compact form of ionization gage detector has been proposed. When high-energy beta (or other) particles from a radioactive source, such as strontium-90, enter a gas, they are able to remove electrons from some of the molecules or atoms of the gas. The resulting gas, which is said to be ionized because it now contains free electrons and residual positive ions (p. 92), is capable of conducting an electric current. If a voltage is applied between two electrodes in the gas, a current, called the ion current, will flow. The magnitude of this current is then a direct measure of the quantity of gas passing through the detector.

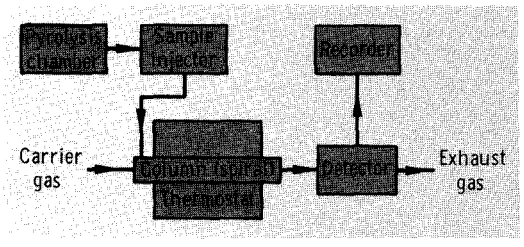


FIGURE 11.1. Schematic representation of gas chromatograph.

A typical output, the gas chromatogram, of such a detector placed at the end of a gas chromatograph column is shown in figure 11.2. It gives the ion-current strength versus time for a particular gas mixture. The time corresponding to each peak is the retention time for a given constituent of the gas mixture. The identity of each of the constituents is determined by comparison with retention times obtained under identical conditions with a mixture of known gases. The area under each peak is a measure of the quantity of the particular gas present in the sample being analyzed. Thus, gas chromatography can provide a complete analysis of a mixture of gases, provided the constituents have been anticipated beforehand.

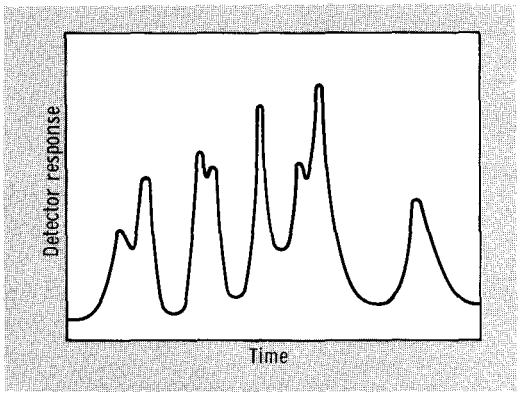


FIGURE 11.2. Typical gas chromatogram.

The way in which it is proposed to use the gas chromatograph to provide some indication of life on Mars is as follows: A sample of the Martian soil is placed in a small oven (pyrolysis chamber), where it is heated electrically to a temperature of 450° to 500° C for about 2 minutes. The organic material and living organisms, if any, present in the soil are decomposed at the high temperature (pyrolysis) and the vapors produced from the

sample are injected into the chromatograph column.

Laboratory tests made on terrestrial soils have shown that proteins, carbohydrates, lipids (fats and related compounds), and possibly nucleic acids give characteristic chromatograms which can be used for identification purposes. It should be possible to detect these substances if they exist in the Martian soil.

Because the composition of the organic material on the surface of Mars is completely unknown, it would probably be desirable to use a gas-chromatographic system of several columns operating under different conditions. Such a system could be designed to detect a variety of substances of biochemical interest, including prebiotic compounds. If the chromatograms obtained should be different from those of pretested terrestrial materials, attempts would be made to duplicate them in the laboratory.

Mass Spectrometer

Like the gas chromatograph, the purpose of the mass spectrometer, which is being developed for possible use on Mars by Klaus Biemann at the Massachusetts Institute of Technology, would be to detect compounds that might be expected to exist or to have existed in living organisms. By means of a mass spectrometer, the masses, or molecular weights, of the various products present in the vapor produced can be determined by heating a sample of organic material, such as may be present on the surface of Mars. From the mass spectrum, i.e., a pattern of the masses, of the pyrolysis products, it is often possible to identify the original material.

In the mass spectrometer, the vapor to be analyzed is first bombarded with electrons emitted from an electrically heated tungsten filament. The molecules present are thereby ionized into electrons and positive ions, and the masses of these ions are determined. The heated filament can serve the dual purpose of vaporizing the solid material under examination and also of ionizing the pyrolysis products. Several different types of mass spectrometers have been developed for various laboratory applications. The description given below may be taken as applicable to a fairly typical instrument, although other designs are in use.

The positive ions first pass between two plates where they are subjected to the action of an electrical field and are thereby accelerated (fig. 11.3). The accelerated ions then pass through a magnetic field where they undergo deflection. The angle of deflection for an ion of given mass depends on the strengths of the electric and magnetic fields. This means that, for a given strength of the two fields, only ions of a particular mass will be deflected through the angle of the instrument and reach the detector (ion collector) at the far end.

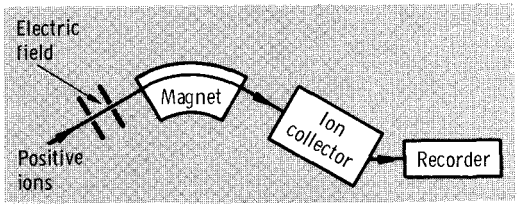


FIGURE 11.3. Schematic representation of a simple mass spectrometer.

By changing the strength of the accelerating electric field and keeping the magnetic field constant, or vice versa, ions of different mass can be detected in turn. The output of the detector, showing how the ion current varies with the strength of the changing field, can be interpreted as a mass spectrum (fig. 11.4). It gives the masses on the conven-

tional atomic weight scale and amounts of the various positive ions present in the sample introduced into the instrument.

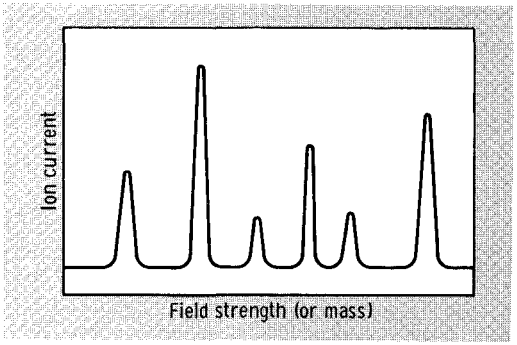


FIGURE 11.4. Typical mass spectrum.

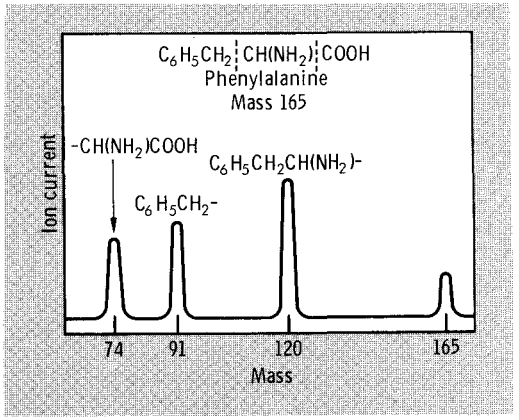


FIGURE 11.5. Mass spectrum of phenylalanine. (After K. Biemann.)

The mass-spectrum technique has been found useful in the laboratory for determining the structures of fairly complex organic molecules. Furthermore, each substance produces a characteristic mass spectrum of pyrolysis products, called a cracking pattern, which can often be used for identification purposes. An illustration of immediate interest is provided by the amino acid phenylalanine, a constituent of many proteins. The mass spectrum of the vapor obtained by heating this amino acid is shown in figure 11.5. It is seen that, in addition to a line corresponding to an ion of mass 165, which is that of the phenylalanine molecule itself, the mass spectrum has lines for three radicals, with mass numbers of 74, 91, and 120, respectively. These are produced by breakage of the molecule at the bonds indicated in the figure. Incidentally, essentially all alpha amino acids (p. 163), such as are found in proteins, yield the radical —CH(NH_2)COOH with a mass of 74 units.

On the surface of Mars, a mass spectrometer with a range of 0 to at least 250 mass units could perhaps be used like a gas chromatograph to indicate the presence of substances of biological significance. The mass spectrometer has an advantage over the gas chromatograph if the material under investigation happens to be one that has not been studied previously in the laboratory. From the masses of the pyrolysis fragments, as indicated by the mass spectrum, information can be obtained which will help in elucidating the nature of the substance.

Gas Chromatograph-Mass Spectrometer

A drawback to the mass spectrometer technique when used in connection with a mixture of compounds of high molecular weight, such as might be found in the Martian soil, is that there are so many lines in the mass spectrum, with overlapping patterns, that identification of individual substances becomes virtually impossible. By combining a mass spectrometer with a gas chromatograph, this difficulty can be largely overcome. The gas chromatograph first separates the pyrolysis products into fractions, according to their retention times, and each fraction is then examined in the mass spectrometer.

The large volume of carrier gas, relative to

the substances extracted in the chromatograph column, results in a decrease in efficiency of the mass spectrometer. Consequently, the gas emerging from the column must pass through a molecular separator which removes most of the carrier gas. One form of separator takes advantage of the larger momentum of the relatively massive pyrolysis products compared with that of the light carrier gas helium. The gas leaving the chromatograph column flows directly toward a disk in which there are holes of small diameter. The pyrolysis products tend to pass straight through the holes to the mass spectrometer, but the lighter molecules of carrier gas diffuse to the side and are pumped away.

Optical Activity: the Pasteur Probe

It has already been mentioned that when an asymmetric substance, capable of exhibiting optical activity, is prepared from nonliving sources, the result is always an inactive racemic mixture consisting of equal amounts of the dextro- and levo-rotatory forms. Asymmetric compounds of biogenic origin, however, are always optically active. Observation of the rotation of the plane of polarization of light would thus provide a decisive indication of the existence of living systems, either at present or in the not too distant past. The qualification, not too remote, in connection with possible past life, is used because there is a natural tendency for all optically active substances to change into the racemic form in the course of time.

Consideration was given originally to the development of a method for testing directly for optical activity of material on the surface of Mars by measurement of the rotation of the plane of polarization of light. This approach was abandoned because the amount of rotation expected would probably be too small to observe with any degree of certainty. Therefore, another, indirect, procedure for detecting the presence of optically active compounds is being studied.

Two optically active forms of an asymmetric molecule, referred to as stereoisomers, have spatial configurations that are mirror images of each other (p. 164). Let the molecule be represented by X; the two stereoisomers would then be D–X and L–X. Suppose X can form a compound (derivative) with another optically active molecule Y. If the L-form of Y interacts with a mixture of D–X and L–X, the compound will be produced in two forms (D–X, L–Y) and (L–X, L–Y), which are not mirror images. Such optically active compounds, having the same chemical formula but different configurations that are not mirror images, are called diastereoisomers. It was found in 1965 by G. Pollock and his associates at the Ames Research Center of the National Aeronautics and Space Administration that the diastereoisomers corresponding to optically active amino acids could be separated by gas-liquid chromatography. This fact is being utilized by J. Lederberg and his coworkers at Stanford University, in a device called the Pasteur probe, to search for optical activity on Mars. The system was named for Louis Pasteur who made the earliest studies on the separation of optical isomers.

The most useful form of the Pasteur probe involves gas chromatography for separating diastereoisomers, followed by a mass spectrometer for their identification. The operation is based on the following general principles: An artificial mixture is prepared of equal molecular proportions of the D- and L-forms of an optically active resolving agent, X. One of the forms, say L–X, differs from the other in the respect that some of its hydrogen atoms have been replaced by atoms of the heavier isotope deuterium (heavy hydrogen). This is indicated by adding an asterisk: L–X*. In the mass spectrometer the

D–X and L–X* residues can be distinguished by the difference in their masses, and the ratio of D–X to L–X* in each peak of a chromatogram can be determined.

Suppose the resolving agent X is combined with an amino acid Y. If the latter is racemic, it will consist of a mixture of D–Y and L–Y, and the combination with the two forms of X will contain four species. When passed through a gas-liquid chromatograph, however, there are only two peaks corresponding to (D–X,L–Y) + (L–X*,D–Y) and (D–X, D–Y) + (L–X*,L–Y). The ratio D–X/L–X* will then be the same in the two peaks. Suppose, though, that the amino acid is present in the optically active from L–Y only. Again there will be two peaks in the chromatogram caused by (D–X,L–Y) and (L–X*,L–Y), respectively. But now, the ratio D–X/L–X* will be quite different in the two peaks; essentially all of the heavier species L–X* will be found in one peak, but very little in the other.

The material proposed as the resolving agent is trifluoroacetyl-thiazolidine-4-carboxylic acid. Both D- and L-forms are available and two of the hydrogen atoms can be easily replaced by deuterium in the L-form to yield L–X*. The chloride of this acid reacts readily with amino acids to form condensation products which are diastereoisomers. Upon heating to a sufficiently high temperature, these condensation products break up into residues corresponding to X and Y. There are actually two residues corresponding to X, with masses of 184 units (D–X) and 186 units (L–X*), and measurement of the signals at 184 and 186 will indicate immediately whether optically active amino acids are present or not. If there are several different amino acids, each will produce a pair of distinct peaks in the chromatogram, and the nature of the acids can be determined from the masses of the Y peaks in the mass spectrogram (fig. 11.8).

A simple application of the Pasteur-probe principle for the detection of metabolism will be described later (p. 239).

Shifts in Dye Spectra

When certain dyestuffs combine with macromolecules of biological origin, including proteins and nucleic acids, the color of the dye changes. The individual components of the macromolecules, such as the simple amino acids, do not produce these changes; a chain containing at least four to six residues appears to be necessary. The color changes are best studied by observing the absorption of light of different wavelengths by the dyestuff alone and also after the addition of the macromolecular substance. The dye alone exhibits absorption maxima at certain wavelengths, but in the presence of the macromolecules new maxima appear at different wavelengths.

A characteristic absorption band of this type, which appears to be produced only by macromolecular species of biological origin, occurs in the vicinity of a wavelength of 6500 Å. This particular band is commonly referred to as the J-band, after E. E. Jelley, who made a detailed study of it in England around 1936. An instrument designed to detect the absorption of light in the region of 6500 Å, called the J-band life detector, was developed by the Aeronutronic Division, Philco Corp. The program has been suspended, however, in favor of others which, for one reason or another, appear to be more promising.

Ultraviolet Spectrum

Every chemical bond between two atoms in a molecule causes an absorption of radiation at a particular wavelength. If the substance has a visible color, the absorption oc-

curs in the visible part of the spectrum, but in colorless compounds the absorption is usually at shorter wavelengths in the ultraviolet region. Of special interest for the problem of determining signs of life on Mars is the absorption band at wavelengths close to 1900 Å (in the ultraviolet) which is characteristic of the peptide linkage (p. 164).

Unfortunately, there are a number of nonpeptide compounds that also absorb ultraviolet light in the same part of the spectrum. It was proposed to distinguish between them and peptides by taking two equal samples of soil from the surface of Mars and extracting each with water in the same way. Both extracts will, of course, give the same amount of absorption in the 1900-Å region of the ultraviolet spectrum. One extract would then be hydrolyzed, that is, broken down into its constituent simple amino acids, by adding hydrochloric or sulfuric acid to the aqueous extract. Because the simple amino acids do not absorb in the region of 1900 Å, a decrease in the absorption, as compared with the unhydrolyzed sample, would indicate the presence of a polypeptide in the original extract. It appears, however, that the treatment with acid may also decrease the ultraviolet absorption of nonpeptide compounds. If so, then the method would be useless. There is, thus, some doubt concerning the value of the ultraviolet absorption spectrum as an unequivocal means of detecting proteins or polypeptides in general.

Detection of Adenosine Triphosphate

It was seen in the preceding chapter that adenosine triphosphate (ATP) is probably found in the cells of all terrestrial organisms, and so its presence on Mars might be a good indication of life on that planet. There is a very simple and sensitive chemical method for determining ATP which simulates the process responsible for light emission by the firefly. It is being developed by Gilbert V. Levin, in an instrument called Diogenes, for possible use on Mars. This is being done in conjunction with the Goddard Space Flight Center of the National Aeronautics and Space Administration. The average light-producing reaction in the firefly apparently involves five substances: a substrate called luciferin, an enzyme known as luciferase, oxygen, magnesium ions, and ATP, in aqueous solution.

If a small quantity of ATP is added to a solution containing luciferin, luciferase, and a magnesium salt, in the presence of oxygen, there is an immediate production of visible light, resembling that of the firefly. Such a solution can serve as a sensitive detector of ATP, and hence of living cells.

Although luciferin is a fairly complex organic compound, it can be synthesized in the laboratory, and the use of optically active D-cysteine in the final stage of the synthesis leads to a product identical with that found in the firefly. Luciferase, however, cannot be synthesized and it is obtained by extraction from firefly tails and subsequent purification.

An interesting aspect of the ATP detector is that, because the response is instantaneous, it is particularly adapted to search for living organisms in the Martian atmosphere. Apart from avoiding the necessity for landing an instrument on the surface of the planet, there is another possible advantage of an ATP detector. An instrument on the surface is immobile and can search for life only in its immediate vicinity. If life on Mars is restricted to a few special locations, as is quite possible, the landed instrument might fail to detect it. Micro-organisms may be widely distributed in the Martian atmosphere, as they are in the terrestrial atmosphere, and so they would be more likely to be observed by an atmospheric probe.

LIFE DETECTION BY METABOLISM

Radioisotope Biochemical Detector

The device called Gulliver, designed by G. V. Levin, then of Hazelton Laboratories, Inc., and N. H. Horowitz, is intended for detection of the metabolism and possibly the growth of Martian organisms, if they exist, by observing the formation of carbon dioxide. On Earth, this gas is a product of the metabolic activities of most life forms and it is assumed that the same might be true on Mars. To provide a means of detecting the carbon dioxide that may be produced in this manner and for distinguishing it from that already in the Martian atmosphere, it is proposed that a nutrient (culture) solution be used containing organic substrates in which some of the carbon atoms have been replaced by the radioactive isotope carbon-14. The carbon dioxide formed by the metabolism of the substrate by a living organism on Mars will then be detectable by its radioactivity.

Among possible substrates which can be readily labeled with carbon-14 are formate, acetate, lactate, glycine, and glucose. The correct choice of substrates is critical to the Gulliver experiment. The ones selected would include materials that are used by terrestrial anaerobic organisms and also carbon compounds that might have been formed in the primitive Martian atmosphere, as indicated by some of the experiments described in chapter IX. Optically active substances would be used in both D- and L-configurations, because it is not known which form might be metabolized by life forms on Mars. In addition to the organic substrates, the culture medium would also contain inorganic salts, such as nitrates, phosphates, etc., which are known to be essential for the metabolism and growth of terrestrial organisms.

In the original form of the Gulliver detector, a sample of soil was drawn, by means of a sticky string, into an incubator (culture) chamber. A sterile nutrient medium contained in an ampule was then admitted by breaking the ampule. A Geiger counter, capable of detecting the beta particles emitted by carbon-14, was attached to the chamber. If metabolism occurred, the formation of radioactive carbon dioxide ($^{14}CO_2$) was indicated by the counter.

A modified version of the Gulliver device weighs only 3 ounces, and consists essentially of an ampule containing the culture solution and a particle counter. The instrument is designed so that when it is expelled from a spacecraft it lands in such a manner as to form a small chamber between itself and the ground. The ampule of solution is then broken and the liquid is distributed on the surface underneath the device. If radioactive carbon dioxide is produced as the result of metabolic activity of micro-organisms in the Martian soil, it can be detected.

Several modifications of the Gulliver principle have been suggested. One is to include in the nutrient medium the amino acid cysteine, a common protein constituent, containing radioactive sulfur-35 in place of its normal sulfur atoms. Metabolism of living organisms would then be indicated by the production of radioactive hydrogen sulfide ($H_2{}^{35}S$) gas, which could be detected by a Geiger counter. In another modification, there is a light source under the instrument. A difference in the rate of formation of radioactive carbon dioxide with the light on and the light off would imply that a photosynthetic process is taking place.

The unique aspect of the Gulliver device is that it uses radioactivity to detect gases that are produced by the metabolism of soil organisms. Although it is a very simple instrument, it has the drawback that it can detect only one gas at a time, and perhaps

not more than two gases, carbon dioxide and hydrogen sulfide, altogether. A proposed development, which would be somewhat more complicated, would be to use a gas chromatograph to detect several gases simultaneously. First, an analysis would be performed on the normal soil gases, and then the changes resulting from the addition of a nutrient solution would be observed.

The Multivator System

The term Multivator, an abbreviation of multiple evaluator, is the name given to what has been described as a miniature laboratory for conducting a variety of biological and biochemical experiments. The apparatus was devised by J. Lederberg and his associates. The essential requirement for the experiments that can be performed is that there is a change which can be detected and measured by means of a photomultiplier tube. The Multivator can consequently be used, in principle, to study biological processes accompanied by such phenomena as fluorescence, turbidity, light scattering, or the production of a substance that absorbs light of a particular wavelength.

The special feature of the Multivator is that it consists of a number of modules, 15 in one design, arranged like the spokes of a wheel (fig. 11.6). In each module, a small lamp is located in front of a light filter which passes light in a narrow range of wavelengths. This is followed by a small reaction chamber, and then another filter (fig. 11.7). For measurements of turbidity or of light scattering, the second filter would pass light of the same wavelength as the first, but for observations of fluorescence or absorption the second filter would transmit at a different appropriate wavelength. The light passing the second filter enters a single photomultiplier tube in the hub of the wheel. Several different experiments of biological interest can then be performed in turn in the miniature laboratory. Not all of the modules are used for experiments, however; some are for control purposes only.

In operating the Multivator on the surface of Mars, an impeller, also located at the hub of the wheel, draws dust into each reaction chamber that already contains the appropriate substrate in dry form. When sufficient dust has been collected, a bellows causes the chamber to close and the solvent to be injected into it. The substrate materials then dissolve and the biological reaction will begin, assuming the required conditions are met. After allowing a predetermined time for the reac-

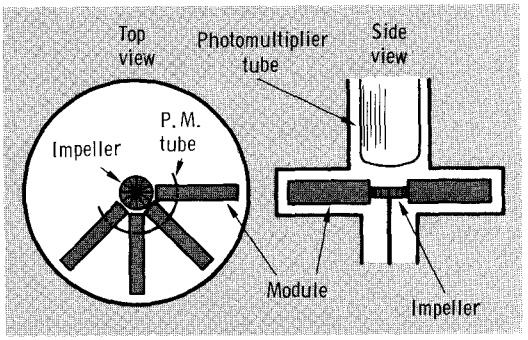


FIGURE 11.6. Representation of top (left) and side (right) views of the Multivator.

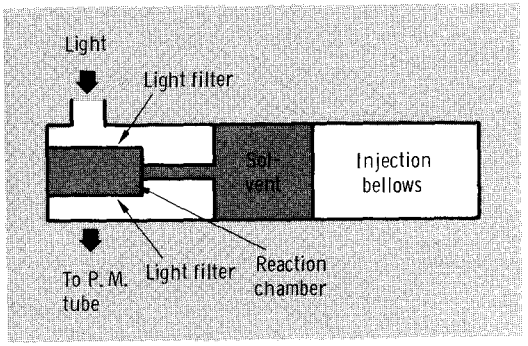


FIGURE 11.7. Construction of a Multivator module.

tion to continue, the lamps in the various modules are switched on in sequence and the outputs are detected in turn by the photomultiplier tube.

One particular biochemical test for which the Multivator is particularly well suited is the determination of the enzyme phosphatase. This enzyme is fairly specific for the hydrolysis of organic phosphates. It is widespread among terrestrial organisms, and because it is involved in the role of the element phosphorus in metabolism and energy transfer, it may be a constituent of Martian life forms if they exist. The dyestuff fluorescein and its derivatives form a phosphate, represented by $X \cdot H_2PO_3$, which is not fluorescent. In the presence of phosphatase as catalyst, it is hydrolyzed (decomposed by water) in aqueous solution to produce fluorescein itself (XH), which is fluorescent, and phosphoric acid (H_3PO_4):

$$\underset{\text{Not fluorescent}}{X \cdot H_2PO_3} + H_2O \xrightarrow{\text{phosphatase}} \underset{\text{Fluorescent}}{XH} + H_3PO_4$$

In testing for the presence of phosphatase in the Martian soil, the substrate would be the fluorescein-phosphate compound and the solvent would be water. The light filter following the reaction chamber would be such that it transmitted only in the wavelength range of the fluorescence produced by the compound XH. Fluorescein fluoresces in the yellow-green region of the spectrum, but some of its derivatives fluoresce at other wavelengths. The production of fluorescence of the appropriate color, as indicated by the photomultiplier tube following the light filter, would imply the presence of the enzyme phosphatase. This would not necessarily mean that living organisms actually exist in the Martian soil, but such organisms must have existed at some time to produce the enzyme.

Oxygen Isotope Exchange

In terrestrial organisms, certain inorganic oxygen-containing anions, that is, negative ions in solution, such as nitrate (NO_3^-), sulfate (SO_4^{2-}), and phosphate (PO_4^{3-}), play an essential role in metabolic processes. In some cases, the anion is reduced and all or part of the oxygen appears as water, whereas in others there is a breakage of bonds between the central atom and one or more of the oxygen atoms in the formation of such substances as adenosine phosphates and phosphosulfates. These reactions in living organisms take place in aqueous solution under the influence of appropriate enzymes.

Oxygen in nature exists as three isotopes, oxygen-16, oxygen-17, and oxygen-18. Of these, oxygen-16 is by far the most abundant, the proportions of the others being relatively small. It is possible, however, to prepare oxygen compounds, such as the anions referred to above, which are enriched in one or the other or both of the isotopes oxygen-17 and -18. That is to say, the proportions of these isotopes, relative to oxygen-16, are greatly in excess of the normal values.

Suppose such an anion, enriched in oxygen-18, for example, is reduced in a living organism in the presence of the appropriate enzyme. The water formed will then contain more than the normal proportion of oxygen-18. Similarly, if the oxygen bonds, such as P—O, are broken in an enzyme reaction, there will be a tendency for the oxygen-18 atoms in the anion to exchange with the oxygen-16 atoms of the common water molecules ($H_2{}^{16}O$) of the aqueous medium. Here again, as a result of the biochemical reaction, the water will contain more of the oxygen-18 isotope than it does normally.

The ability to catalyze the change in the isotopic ratio, that is, to cause an increase in the ratio of oxygen-18 to oxygen-16, in water

containing certain anions appears to be unique to enzymes found in living organisms. The possibility of using this situation as a test for life on Mars is being investigated by B. Kok, J. E. Varner, and others at the Research Institute for Advanced Studies, Baltimore.

The ratio of oxygen-18 to oxygen-16 can be determined by means of a mass spectrometer similar to the instrument described earlier. For the measurement, a sample of water vapor may be admitted to the mass spectrometer. Alternatively, a small quantity of liquid water may be decomposed by electrolysis and the oxygen evolved can be used for analysis. A different approach is to enrich the anions in oxygen-17, instead of oxygen-18. A change in the proportion of oxygen-17 in the water can then be measured by the nuclear magnetic resonance technique. This depends on the fact that oxygen-17 nuclei, unlike the nuclei of oxygen-16 and oxygen-18, behave as small magnets which can be alined in a magnetic field.

Change in Optically Active Substrate

A relatively simple method for detecting metabolism is based on the separation of diastereoisomers by gas chromatography, as mentioned earlier in the description of the Pasteur probe. If a living terrestrial organism is supplied with the D- and L-configurations of a given substrate, such as an amino acid, it will consume one form, the L-amino acid, for example, and leave the other form. In this case the metabolism can be detected by a change in the ratio of the D-form to the L-form.

To determine any change in the D/L ratio, the amino acid, which may be represented by Y, is combined with an optically active form, say L–X, of a suitable resolving agent. The system will then contain the diastereoisomers (L–X, D–Y) and (L–X, L–Y), which give separate peaks in a gas chromatogram. For the detection of life on Mars, a sample of soil would be added to water containing the D- and L-forms of a mixture of amino acids. After the lapse of a few hours, some of the clear liquid would be removed, dried, and the residue converted into diastereoisomers. The product would then be subjected to gas chromatography in the usual way. If there are living organisms on Mars, and if, like organisms on Earth, they have a preference for one molecular configuration of amino acids over the other, then the relative heights of the peaks in the chromatogram would change.

An illustration of the results obtained in this manner with terrestrial soil and a mixture of D- and L-forms of three amino acids—glutamic acid, proline, and valine—is given in figure 11.8. The gas chromatogram at the left was obtained before the soil was added to the amino acids and the one on the right was recorded 48 hours later. It is clear that, in each case, the ratio of L-amino acid to the D-configuration has decreased because of the consumption of the former by the micro-organisms in the soil.

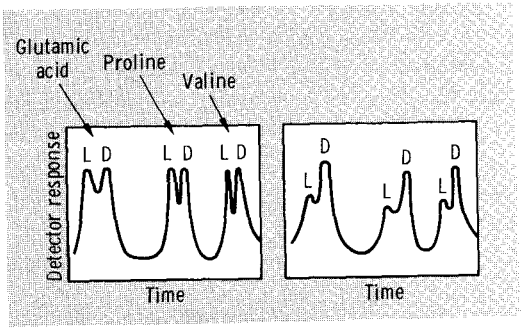


FIGURE 11.8. Gas chromatograms of mixture of amino-acid diastereoisomers before (left) and after (right) addition of terrestrial soil. (After B. Halpern, J. W. Westley, E. C. Levinthal, and J. Lederberg.)

LIFE DETECTION BY GROWTH AND MORPHOLOGY

The Wolf Trap

The instrument called the Wolf trap, so named for its inventor Wolf Vishniac (p. 213), is designed to detect the growth and metabolism of micro-organisms by measurement of light scattering and the change in hydrogen-ion concentration; that is, in the pH. When terrestrial organisms grow in a clear nutrient solution, the liquid soon becomes turbid and it is assumed that Martian organisms would produce the same effect. The development of turbidity could be detected by attenuation of a beam of light passing through the medium, but it appears that observation of the light scattered out of the beam would be a more sensitive indication of the growth of organisms. Growth and metabolism of terrestrial organisms is also accompanied by a change in the acidity (pH), and this can be readily determined by means of suitable electrodes inserted in the solution.

In testing for life on the surface of Mars, a sample of soil would be drawn into an incubation chamber containing the nutrient solution. Any light scattering and pH change would then be recorded over a period of time. If growth of micro-organisms occurs, the rate could be estimated from the way in which the scattering of light increases with time. In the event that there are no organisms present or those that are present do not grow, there would be no light scattering or change in pH.

There are some factors which may interfere with the proper operation of the Wolf trap device. If the sample from the surface of Mars introduced into the culture chamber contains large amounts of colloidal material, as terrestrial soils generally do, the resulting scattering might obscure that caused by the growth of micro-organisms. On the other hand, certain colloids swell when suspended in water and could thus cause an increase in light scattering even in the absence of growth. Furthermore, the pH changes or absence of them might be misleading. If the Martian soil contains relatively insoluble acidic or basic substances, the pH of the medium will change slowly in the course of time as the substance dissolves, even if there is no growth of micro-organisms. If the soil should be highly buffered, as is often the case on Earth, the pH might not change noticeably although micro-organisms were growing in the nutrient solution.

Because the increase in the scattering of light by a liquid medium provides a useful indication of the growth of micro-organisms, E. L. Merek and V. I. Oyama, at the Ames Research Center, have developed a procedure which avoids the interference arising from the presence of colloidal materials. The culture chamber is separated into two parts, upper and lower, by means of a sintered glass plate. A dry sample of soil is introduced into the upper section and the liquid nutrient medium into the lower one. As a result of capillary action, the liquid rises in the sintered glass plate and wets the soil. Micro-organisms can then migrate and grow from the soil into the medium in the lower chamber, which thus becomes turbid. Although the organisms can pass through the sintered glass, the colloidal particles do not, and so the difficulties mentioned above do not arise.

Calorimetry

All terrestrial living organisms are characterized by the evolution of heat as a consequence of their metabolic processes. If reproduction occurs and the number of organisms increases, the rate of heat release increases exponentially with time. If there are

forms of life on Mars, it is reasonably certain that they will behave in a similar manner. Moreover, the evolution of heat should occur regardless of the biochemistry of the Martian organisms, whether similar to that of terrestrial organisms, as is postulated in some other tests for life on Mars, or not. An advantage of the calorimetric method is that it would not be necessary to use a special nutrient medium, because the Martian soil itself could be used for this purpose.

The possibility of utilizing a microcalorimeter to detect the growth of micro-organisms on Mars was studied by Beckman Instrument, Inc. It was decided that the technique was not sensitive enough for the intended purpose, and the studies have been discontinued, at least for the present.

The Vidicon Microscope

A simple and often effective way of studying the reproduction, as well as the morphology, of terrestrial micro-organisms, is to observe them in a microscope. The use of a microscope combined with a television-type (vidicon) camera was suggested by J. Lederberg and is being studied at the Jet Propulsion Laboratory. The vidicon microscope is a phase-contrast instrument with a fixed focus; this means the positions of lamp, focal plane, and objective lens are fixed relative to one another (fig. 11.9). The image of the specimen under observation is formed on the face plate of a vidicon tube. This plate is coated with a material which becomes an electrical conductor on those parts that have been exposed to light. The conductivity is retained for a sufficient time to permit scanning by an electron beam, as in a television camera. The microscope image can thus be transmitted to Earth by telemetry.

To introduce a sample of Martian soil, an aerosol, consisting of fine particles suspended in a gas, is injected in front of the microscope condenser lens, as indicated in figure 11.9. The suspension then passes through a hole in the condenser and the particles are deposited on a transparent impaction plate located at the focal plane of the microscope.

Another type of instrument under development is the automated scanning, flying-spot photometric microscope, which has been used by the U.S. Army Biological Laboratories to distinguish living organisms from dust particles. An aerosol produced from the surface of Mars would be allowed to impact on, and adhere to, clear plastic tape. After passing through a staining solution, which stains the organisms but leaves the inorganic dust particles unaffected, the tape is scanned by means of a narrow moving beam (flying spot) of light. The differences in transmission of light as the flying spot encounters stained living organisms and unstained, generally opaque, inorganic particles can be recorded.

The Electron Microscope

The electron microscope is normally a large and massive instrument unsuitable for transport to Mars by a spacecraft, but H. Fernández-Morán, of the University of Chicago, an authority on electron microscopes, is

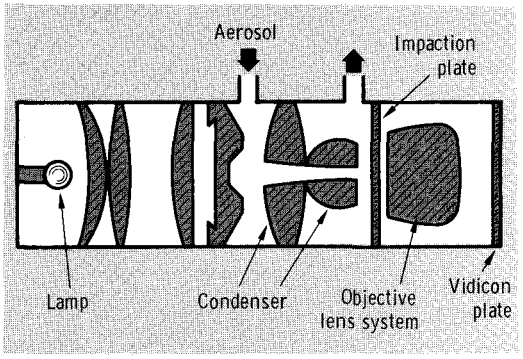


FIGURE 11.9. A vidicon microscope.

of the opinion that the size and weight could be greatly decreased. Such an instrument would have a resolving power at least 10 times that of the best optical microscope and it would be able to detect objects less than one-tenth of the size of the smallest that can be seen in a conventional microscope. This would represent a great advantage from the morphological standpoint, because it would be possible to examine the internal structure of any small particles suspected of being living organisms.

THE AUTOMATED BIOLOGICAL LABORATORY

The Laboratory Concept

Because life on Mars, if it exists, is uncertain and unpredictable, it will be necessary to perform many different tests in the areas of chemistry, metabolism, reproduction, and morphology. If such tests were made on Earth, they would be carried out in a logical manner, with the results of each test determining the nature of the succeeding one. The possibility of doing something similar on Mars by remote and computer control has led to the concept of the automated biological laboratory for landing on Mars.

A report prepared in 1965 for the National Aeronautics and Space Administration states:

> In the automated biological laboratory concept, the experimental program is conducted, not with individually mechanized experiments, but by an integrated complement of basic instruments operated in a sequential fashion . . . [just as] biological experimentation is performed in a terrestrial laboratory. The [automated] laboratory is controlled by an on-board computer, with command override capability provided for Earth-based scientists to select alternative experimental programs, or even to initiate completely new programs in response to the results obtained from preceding experiments. . . .
>
> In addition to life detection experiments, it is useful to include environmental experiments which will provide information for use in the performance and interpretation of the biological investigation. The inclusion of . . . [these] experiments also provides capability . . . [for the performance] of characterizations of other kinds . . . , such as geological, meteorological, and paleontological. Such capability makes it possible to obtain useful information even if selected life detection experiments produce negative results. For these reasons, a comprehensive complement of both life detection and environmental experiments should form the basis of any investigation to detect life on Mars.

In an effort to implement the foregoing requirements, the Aeronutronic Division of the Philco Corp., which was responsible for this report, has prepared a detailed design for a proposed automated biological laboratory. The package to be landed on Mars, including a self-contained power source, would have a roughly spherical form with a diameter of almost 6 feet and an estimated weight of about 1200 pounds (fig. 11.10). It would have the capability of performing up to 35 experiments. Of these, 7 are concerned with the Martian atmosphere and environment, 10 with the physical and chemical properties of the soil, and the remainder with various tests, similar to those described earlier in this chapter, for studying the characteristics of living systems. Since the automated biological laboratory will probably not become a reality for several years, the details of the design and experiments are subject to change. The subsequent discussion will, therefore, be confined to general principles.

Observations in the Atmosphere

Measurements on the Martian atmosphere, made by an automated biological laboratory, would include determination of the temperature, pressure, and composition. The temperatures at the surface and at distances up to about 6 feet above the surface, and their diurnal variations, could be observed. Measurements of pressure would include both static and dynamic values. The

static pressure is the normal barometric pressure, whereas the dynamic pressure is that produced by winds. In addition to determining the main constituents of the Martian atmosphere, which are thought to be carbon dioxide, nitrogen, and argon, a combination of gas chromatograph and mass spectrometer would indicate the presence and amounts of any oxygen and water vapor, as well as of other gases of biological and geological interest, such as methane, hydrogen sulfide, oxides of nitrogen, sulfur dioxide, and so on. Because of the great importance of water for life processes, an especially sensitive detector has been developed to measure the amount of water vapor in the atmosphere (p. 283).

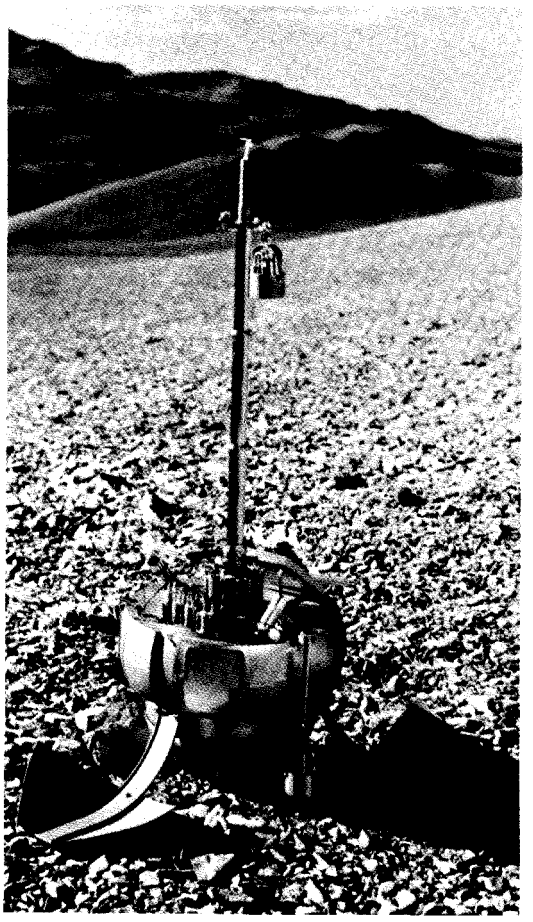


FIGURE 11.10. Model of proposed Automated Biological Laboratory.

Studies of the wind and its variations throughout the Martian day and night would be made from the automated biological laboratory. The velocity of the surface wind and its direction are significant in connection with the possibility that life forms have become adapted to removal of or addition to their soil cover. Winds may also have a bearing on the distribution of micro-organisms over the surface of the planet. The ability of the wind to carry small particles is related to the viscosity and density of the atmosphere. These quantities can be calculated from the temperature, pressure, and composition. Airborne particles and dust could also be collected and examined for size distribution, chemical composition, and possibly for the presence of living organisms.

Other meteorological observations would include scanning of the sky in both visible and infrared regions of the spectrum. This would provide information on cloud cover, haze, duststorms, sky brightness, and transmissivity of the atmosphere. These factors may affect living organisms through their influence on the heat balance of the planet and the large-scale transportation of dust, spores, and other particles, and also of water vapor.

Various radiations from the Sun and from space may determine the conditions under which life could exist on Mars. Instruments would determine the intensity of solar radiation in the ultraviolet, visible, and infrared regions. The ultraviolet irradiance, in the range of 1850 to 3000 Å, would be of special interest because its magnitude, which could be critical, may be affected by the circumstances responsible for the blue haze.

Measurements would also be made of the

primary cosmic rays, both from the Sun and from beyond the solar system, consisting mainly of high-energy protons, and of the secondary particles produced by their interaction with the Martian atmosphere. On Earth, the latter are far more significant than the primary cosmic-ray particles. The data would provide much needed information on the possible effects of the radiations on living organisms; for example, in producing mutations. The beta and gamma rays from radioactive elements, especially uranium, thorium, and potassium in the soil, are in the same category as cosmic rays with regard to their biological effects. These radiations in the atmosphere would also be measured.

Studies of the Martian Soil

The next series of experiments performed with the instruments on the Automated Biological Laboratory are concerned with the physical properties of the Martian soil. Measurements would be made of the temperature at several levels at and below the surface, and of its daily variations. From simultaneous determinations of the amount of water vapor, it would be possible to derive information about the condition of any water in the soil. Such water may conceivably be present as liquid or solid, or bound, chemically or physically, by soil materials. Important characteristics of soil are its density and the sizes of the particles present. These could be determined in various ways. Rough indications of particle-size distribution could be obtained by sieving a soil sample, but microscopic examination would be preferable.

Further information on the structure of the soil can be derived from simple mechanical tests made to determine the strength and penetrability of the surface. Incidentally, this information would be useful in selecting the best method for collecting samples of soil from and below the surface. Nearly all the measurements described so far have not required samples, but they are necessary for most of the subsequent tests.

The ideal Automated Biological Laboratory would have a certain degree of mobility, so that, with the aid of its television cameras, it could be directed to what appeared to be a promising location. The earlier models, however, will undoubtedly be stationary. Nevertheless, there may still be some choice available, at the direction of a controller on Earth, concerning the area in the vicinity of the laboratory from which samples are taken. The sampling devices will include a drill, for obtaining specimens of soil below the surface, as well as, possibly, scoops, scrapers, brushes, suction apparatus, and so on. After a sample is taken it would be passed through a series of sieves, and the smaller particles, which are probably of major interest, would be subjected to various tests.

In continuing the study of the Martian soil, samples would be examined for their chemical characteristics. Perhaps the most important of these is the presence or absence of organic carbon. If there are no organic compounds in the soil, the prospects of finding living organisms would be very poor. If the tests indicate that organic compounds are present, the next step would be analysis for the actual substances by means of a gas chromatograph-mass spectrometer system. The soil would first be heated to somewhat above 100° C, to determine the water content and the nature and amounts of contained gases. Pyrolysis at higher temperatures would then be used to identify organic compounds of various types.

The total elemental analysis of the soil and the determination of soluble ions present are of biological significance. The elements of principal interest are considered to be carbon, nitrogen, oxygen, magnesium, silicon, phos-

phorus, sulfur, potassium, calcium, and iron. Most of these could be determined by the alpha-particle scattering technique used on the Surveyor spacecraft in 1967 and 1968 to analyze the material on the surface of the Moon (p. 284). For characterization of soluble inorganic ions, such as those of sodium, magnesium, potassium, and iron (positive ions), and nitrate, phosphate, sulfate, and chloride (negative ions), which are important for terrestrial organisms, a sample of fine soil would be extracted with water. The resulting solution or the solid remaining after evaporating off the water would then be analyzed by suitable spectrographic, colorimetric, and other methods.

If there is biological activity in the soil, the composition of the entrained gas will be significantly different from the gases in the normal atmosphere. Consequently, analysis of the gas that can be sucked off from a sample of soil would be of interest. If possible, the soil gas should be obtained from various depths below the surface. The components of major importance are methane, carbon monoxide and dioxide, oxygen, ozone, hydrogen sulfide, hydrogen, ammonia, nitrogen, oxides of nitrogen, and sulfur dioxide. Water vapor is not included in this list because it would probably be determined separately. The gas analysis would be performed by a combination of chromatograph and mass spectrometer.

Biological Experiments

Some of the tests described above made on the Martian soil might provide indications of biological activity. In addition, the Automated Biological Laboratory would include equipment for performing various experiments relating specifically to life processes. These would be designed to detect chemical compounds and properties, metabolism, reproduction, and morphology that are considered to be characteristic of living organisms. A number of possible experiments along these lines have already been described, and a selection would be made from these and others which may be developed in due course.

One of the problems associated with experiments intended for the detection of metabolism and reproduction of living organisms is the selection of a nutrient medium containing suitable substrates. Because nothing is known about the postulated life on Mars, it is not possible to do more than make guesses at the nature of the substances, based on experience with terrestrial micro-organisms. A way out of this difficulty has been proposed for use on the Automated Biological Laboratory. Samples of the Martian soil would be extracted with a quantity of water until a moderately concentrated solution had been produced. If there are organisms present in the soil, they would certainly be expected to survive and grow in such a culture medium.

Conclusion

If a significant number of tests indicate that forms of life exist on Mars, it will be essential to determine if they are indigenous to that planet or if they have evolved from terrestrial organisms transferred in some manner to Mars. A resolution of this situation may prove to be very difficult and may require the return of samples to Earth for detailed examination. Of course, if the amino acids on Mars prove to have the D-configuration or sugars the L-configuration, there will be little doubt that life on Mars and on Earth originated independently. Otherwise, it would be necessary to determine if the amino acids and other molecular units, such as pyrimidine and purine bases, present in the Martian proteins and nucleic acids, assuming they exist, are the same as or different from those of terrestrial

organisms. Even if they are the same, however, the two life forms may conceivably have developed independently but along identical lines. Here is a problem which may provide a challenge for biologists in the 1970's.

Suppose, on the other hand, that most, or all, of the tests give negative results, with no definite indication of the presence of life on the surface of Mars. A possible explanation would be that the instruments have landed in an area where there are no living organisms, but life forms may nevertheless exist at other locations. The concept of microenvironments on Mars, where local conditions might be favorable for life, has been considered by a number of scientists. Such favorable microenvironments might occur where volcanic heat has caused the subsurface permafrost to melt and produce liquid water. The positions of these possible regions are completely unknown.

It was thought at one time that life on Mars would most probably be found on areas near the equator, such as Syrtis Major, which remain quite dark in color throughout most of the year. This belief was based on the view long held that the dark regions were low-lying areas and the seasonal darkening was associated with the presence of additional water vapor in the atmosphere. It is possible, however, that the dark portions of Mars are elevated regions and the darkening effect is caused by the removal of fine particles by the seasonal winds. If this is so, then the dark areas might be unsuitable for the development of life.

The search for life on Mars is obviously a quest fraught with difficulties. There is no reason, however, for not pressing the search. If it can be established that living organisms originated on Mars independently of life on Earth, the scientific significance will greatly transcend all the difficulties and problems.

In any event, if there is at present no life on Mars, the search may still provide valuable biological information. As stated at the beginning of this chapter, evidence for the remains of extinct life or for the presence of prebiotic substances would be of great interest. The experiments described here are designed to throw light on various phases of biological evolution. If the results are applicable to any one or more of these phases, the effort will have been worthwhile.

FOR FURTHER INFORMATION

BOTAN, E. A.: An Instrumental Search for Extraterrestrial Life. Space Sci. Rev., vol. 3, 1964, p. 715.

HALPERN, B.; ET AL.: The Pasteur Probe: An Assay for Molecular Asymmetry. Vol. V of Life Sci. Space Res., North-Holland Publ. Co., 1967, p. 239.

KOK, B.; AND VARNER, J. E.: Extraterrestrial Life Detection by Means of Isotopic Oxygen Exchange. Vol. V of Life Sci. Space Res., North-Holland Publ. Co., 1967, p. 206.

KOK, B.; AND VARNER, J. E.: Extraterrestrial Life Detection Based on Isotopic Oxygen Exchange Reactions. Science, vol. 155, 1967, p. 1110.

MAMIKUNIAN, G.; AND BRIGGS, M. H., EDS.: Current Aspects of Exobiology. Pergamon Press, 1965.

PITTENDRIGH, C. S.; VISHNIAC, W.; AND PEARMAN, J. P. T., EDS.: Biology and the Exploration of Mars. Natl. Acad. Sci.-Natl. Res. Council Publ. 1296, 1966.

QUIMBY, F. H.: Some Criteria of Living Systems Useful in the Search for Extraterrestrial Life. Develop. Ind. Microbiol., vol. 5, 1964, p. 224.

QUIMBY, F. H.: Concepts for Detection of Extraterrestrial Life. NASA SP–56, 1964.

The Search for Extraterrestrial Life. Adv. Astronaut. Sci., vol. 22, 1967.

YOUNG, R. S.; PAINTER, R. B.; AND JOHNSON, R. D.: An Analysis of the Extraterrestrial Life Detection Problems. NASA SP–75, 1965.

YOUNG, R. S.: Automatic Life Detection. Astronaut. Aeron., vol. 3, 1965, p. 370.

XII

Avoidance of Contamination of Mars

THE COSPAR SPECIFICATIONS

The Importance of Sterilization

IF THE SEARCH FOR LIFE FORMS on Mars is to have any significance, it is obvious that all conceivable steps must be taken to minimize the possibility that the planet will become contaminated by terrestrial organisms. Micro-organisms in large numbers are found almost everywhere on Earth. They will, therefore, inevitably be present on the spacecraft and instruments to be used for the exploration of Mars. Unless special precautions are taken to destroy them, terrestrial organisms might thus be introduced on the planet.

Essentially all the tests described in chapter XI, proposed to seek evidence of life on Mars, are designed to be applicable to microscopic forms of life such as might be present in soil. The effects of contamination by micro-organisms from Earth would thus be particularly serious. The possibility might arise that an instrument package landed on Mars for the detection of life would provide a source of contamination for its own experiments. The organisms detected would then not necessarily be life forms indigenous to Mars, but those carried to the planet from Earth. If the terrestrial organisms remained localized, the situation would not be as serious as it would if they spread from the point of original contamination and multiplied on the surface of Mars. These terrestrial organisms might then be detected at a later date by a completely sterilized space vehicle and be mistaken for Martian organisms. There is also the possibility that organisms introduced from Earth would affect, and perhaps destroy, the indigenous life forms. It is necessary, therefore, that special efforts be made to prevent the contamination of Mars by terrestrial micro-organisms.

In the volume entitled "Biology and the Exploration of Mars," published by the National Academy of Sciences-National Research Council in 1966, the American biologist K. C. Atwood wrote as follows:

> It is possible that Mars may provide our only chance to study life forms of an entirely separate descent. The importance of this opportunity outranks every prospect; the effect on world thought of the discoveries to be made on Mars may compare with that of the works of Copernicus or Darwin. A careless blunder would destroy the opportunity. The reasons for the sterilization of spacecraft are clear: we do not know enough about Mars to predict with confidence the outcome of

319-889 O—69——8

microbial contamination of its surface. Regrettable and irreversible results can be imagined. The means of avoiding these results are known; hence we must employ such means.

The importance of sterilizing spacecraft designed for the exploration of the planets was realized in the United States at the time the first satellites were placed in orbit about Earth. In February 1958, the Council of the National Academy of Sciences expressed concern that early landings on the planets "might compromise and make impossible forever critical scientific experiments" and agreed to "plan lunar and planetary experiments . . . so as to prevent contamination of celestial objects in a way that would not impair the unique . . . scientific opportunities." As a result of resolutions passed by the Council, the International Council of Scientific Unions in March 1958 appointed a Committee on Contamination by Extraterrestrial Exploration to consider all aspects of the problems associated with the sterilization of spacecraft.

In October 1958, the Committee on Space Research, commonly known as COSPAR, was established by the International Council of Scientific Unions. Its purpose was to provide for the continued and expanded international cooperation in space science which had been initiated so successfully during the International Geophysical Year of 1957–58. In March 1959, COSPAR accepted the responsibilities originally assigned to the Committee mentioned in the preceding paragraph. Subsequently, a Study Group on Standards for Space Probe Sterilization, with members from the United States, the U.S.S.R., the United Kingdom, Sweden, Belgium, and France, was appointed by COSPAR. The recommendations made by this Group have been accepted by the members of the International Council of Scientific Unions. The two countries primarily concerned, however, are the United States and the U.S.S.R.

Planetary Missions and Contamination Modes

Before considering the COSPAR quarantine recommendations, it is necessary by way of introduction to discuss the types of space missions proposed for the exploration of the planets. The simplest mission, called a flyby, involves a spacecraft launched from Earth in an orbit about the Sun so that it passes close to a particular planet. The Mariner IV mission of 1965, for example, was a flyby of Mars (ch. XIII), and the Mars-Mariner missions of 1969 are to be flybys. Normally, in a flyby the spacecraft would be guided to pass within not less than about 2000 kilometers (1250 miles) of the planet's surface.

A second type of mission would include a landing vehicle or lander, referred to as a "capsule," containing instruments for making various observations on the surface of the planet. The landing capsule is carried to the planet on a "bus" and after the capsule is released, the bus may continue on past the planet as a flyby in a solar orbit.

Another possibility is for the spacecraft, or bus, to go into an orbit around the planet. It is then called an orbiting spacecraft or orbiter. Finally, the second and third types of mission may be combined: a capsule is landed onto the surface of the planet, whereas the bus goes into planetary orbit.

Of the mission components described above, only the capsule would be sterilized. The bus carrying the capsule, whether it was intended to go into orbit or not, would not be sterilized. But it would probably be fabricated under such conditions as to keep the microbial contamination to a minimum. In the course of its journey to the planet, the sterilized capsule would be separated from the nonsterile bus by a barrier impervious to micro-organisms. A space vehicle to be placed in orbit or designed to fly by the planet, and which does not carry a capsule, would be

treated like a bus in that it would not be sterilized, but contamination would be kept as small as reasonably possible.

The various ways whereby the surface of the planet could become contaminated are summarized in figure 12.1. The sources of contamination may be divided into three main categories. These are, first, the presence of some micro-organisms in a capsule that was supposed to be sterile; second, the accidental impact of an orbiter or flyby; and, third, other sources, essentially materials that might be accidentally ejected or emitted from various parts of a nonsterile spacecraft. It may be mentioned in connection with the second of these three items that, in the course of time, the path of an orbiter will inevitably decay and the spacecraft will eventually crash into the surface of the planet. But current plans call for the decay period to be so long that, by the time the orbiter reached the surface, the planet would be contaminated in other ways; for example, by manned landings. An error in the trajectory of an orbiter or a flyby, however, could lead to an accidental impact.

On the basis of considerations that will be outlined below, COSPAR adopted in 1964 the recommendations made by the Study Group on Standards for Space Probe Sterilization. These included specifications for the

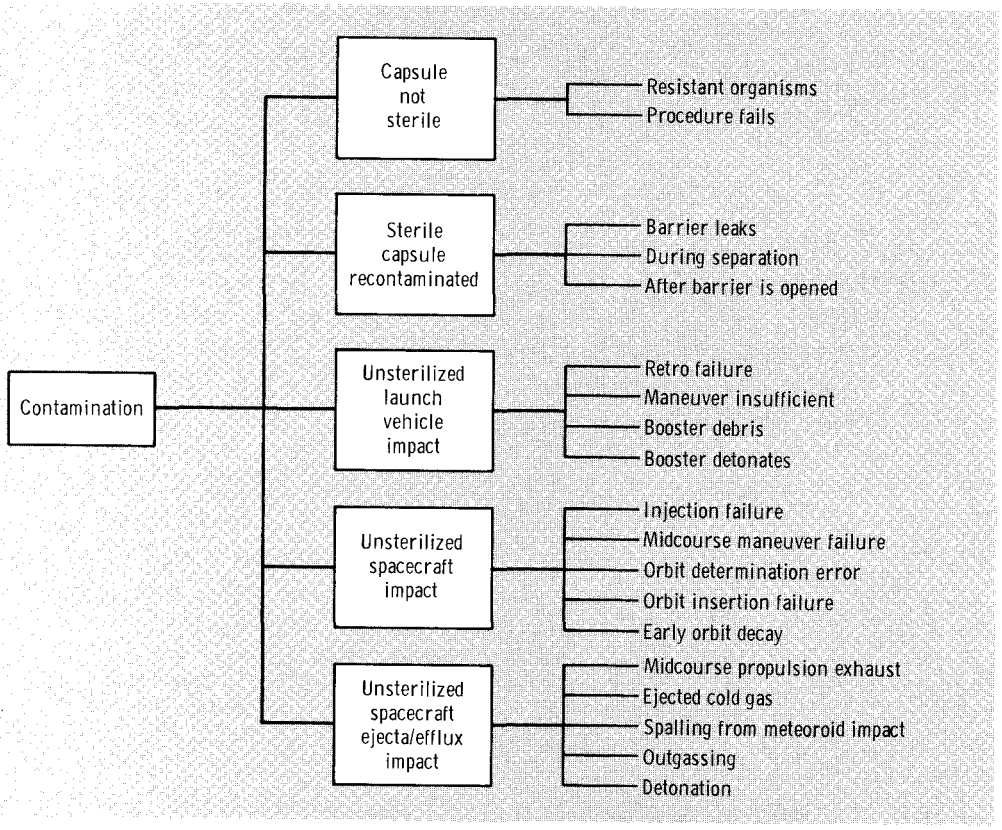


FIGURE 12.1. Possible modes of contamination of the Martian surface by spacecraft.

sterilization of landing vehicles and for the probability of accidental impact of any unsterilized spacecraft. It was realized that the proposed severe quarantine constraints were to be applicable to the early stages of planetary exploration, while the planet was still presumably free from terrestrial organisms. Later, when a reasonable number of experiments had been made in the search for life, the specifications would be relaxed. When men land on the surface of a planet, contamination will be much more difficult to control than it is with unmanned spacecraft.

The COSPAR recommendations were consequently made for all missions "during the interval terminating at the initial period of planetary exploration by landing vehicles (approximately one decade)." The estimate of about 10 years for the "initial period," presumably that elapsing between the first unmanned and the first manned landings, may be too short. In effect, however, the specifications allow for a longer period, as will be apparent.

The COSPAR resolution, which was accepted at the annual meeting in Florence, Italy, in May 1964, included "as tentatively recommended interim objectives, a sterilization level such that the probability of a single viable organism aboard any vehicle intended for planetary landing . . . would be less than 10^{-4}, and a probability limit for accidental impact by unsterilized flyby or orbiting spacecraft of 3×10^{-5} or less." The specification for landers implies that the capsule should be sterilized to such an extent that there is not more than one chance in 10 000 of a single organism capable of reproduction remaining on it. For flyby or orbiting spacecraft, the implication is that there is less than one chance in about 33 000 of an accidental crash on the surface of the planet. These requirements are clearly very severe, but they did not seem to be completely unattainable.

In 1966, the COSPAR Consultative Group on the Potentially Harmful Effects of Space Experiments recommended that one of the quantities included in the assessment of the probability of less than 10^{-4} for a single viable organism aboard a sterilized lander be decreased to one-tenth its previous value (p. 254). This change would have the effect of increasing the probability of 10^{-4} by a factor of 10 to 10^{-3}. The Consultative Group stated, however, that the change "should . . . not . . . be incorporated into the over-all probability until the general form[ul]ation has been agreed upon." This formulation depends on the development of a detailed model of planetary contamination by space vehicles to which reference will be made later. In the meantime, it is commonly assumed, although it is not a COSPAR recommendation, that the probability of a single viable organism remaining aboard a capsule for landing on Mars should be kept below 10^{-3}.

Basis for COSPAR Specifications

One of the first definite proposals concerning the degree of sterilization to be required for a capsule landing on the surface of a planet was made by R. W. Davies and M. C. Comuntzis in 1959. They suggested more or less arbitrarily that the maximum risk of contaminating a planet be kept below 10^{-6} per mission. No effort was made at the time, however, to translate this into the number of organisms per spacecraft. Around 1962, however, a general feeling was voiced, as in a "Review of Space Science," issued by the National Academy of Sciences-National Research Council, that a Mars capsule be sterilized so that "there is less than a 10^{-4} probability that a single living organism will be released on the planet."

Some justification for this particular choice

was published in 1963 by L. D. Jaffe of the Jet Propulsion Laboratory:

> For engineering use, it is necessary to prescribe, numerically . . . the risk of contamination that will be taken. (It is not possible to take zero risk, except by abandoning the exploration of Mars.) One way to obtain a reasonable number is to say that we should keep the chance of contamination of Mars as low as the chance that we will obtain no useful biological data for other reasons.

For the purpose of his calculation, Jaffe postulated a period terminating in 1980 during which efforts would be made to land unmanned capsules on the surface of Mars. In this period there would be seven Martian oppositions (p. 39), at each of which the United States and the U.S.S.R. would each attempt two flights to Mars. There would then be a total of 28 attempted flights. For each attempt, the probability of reaching the planet was assumed to be 50 percent; that is, 0.5. Even if the spacecraft lands on Mars, there may be failures in the landing procedure, power supply, radio communication, telemetry, and so on. The probability of such failure was also set at 50 percent of the successful landings, or 0.25 of the total flights attempted.

The probability that any particular flight will fail to provide useful data is $0.5+0.25=0.75$, and the total probability that no useful data will be obtained in the 28 attempted flights is $(0.75)^{28}=3.2\times10^{-4}$. Because only half of these flights are assumed to reach Mars, the probability of obtaining no useful data per flight may be taken as $(3.2\times10^{-4})/14=2.3\times10^{-5}$.

On the basis of approximate, but somewhat arbitrary, engineering estimates, the probability that microbes would be released from an unmanned spacecraft because of a crash landing was set at 10^{-2} during the whole program. The probability of contamination per flight would then be $10^{-2}/14=7\times10^{-4}$. As stated above, Jaffe suggested that the chance of contamination should be as low as the chance that no useful information will be obtained. From a comparison of the respective values of 7×10^{-4} and 2.3×10^{-5}, he concluded that "perhaps an intermediate value of 10^{-4} . . . is reasonable," for the probability of contamination per flight. Thus, he argued, for a Mars lander "the factor of 10^{-4} becomes the permissible likelihood that a viable organism is aboard."

Although this number is the same as that in the specification recommended for landers by COSPAR in 1964, the latter was based on a different, more detailed, approach to the problem made by C. Sagan and S. Coleman. The discussion of this approach to be given here is simpler and less rigorous than the original, but it leads to essentially the same conclusions.[1] Let σ be the mean number of viable micro-organisms per capsule that are distributed outside the capsule on the Martian surface, and let $P(g)$ be the probability that such an organism will contaminate a significant area of the planet during the course of the total experimental program of unmanned landers. Provided σ is small, the mean probability of significant contamination per capsule is then $\sigma P(g)$, and the probability that there will be no contamination is $1-\sigma P(g)$ per capsule.

Suppose there are N missions in which capsules are landed successfully on the surface of Mars during the course of the experimental program. If it is assumed that the probability of contamination, or of noncontamination, per mission is always the same as the average value derived above, the total probability $P(NC)$ of noncontamination in the N landing missions is then given by

$$P(NC)=[1-\sigma P(g)]^N$$

[1] Some of the symbols have also been changed to comply with a later, more general, usage.

It is an important objective of the program to make σ, the number of viable organisms reaching the Martian surface, as small as possible; hence, $\sigma P(g)$ will be very much less than unity. The expression just derived for $P(NC)$ can then be represented by the first two terms of a binomial expansion, that is

$$P(NC) = 1 - N\sigma P(g)$$

This is the probability of noncontamination, and the corresponding probability of contamination, $P(C)$, which is equal to $1-P(NC)$, is then

$$P(C) = N\sigma P(g) \qquad (1)$$

for capsules landing on the surface of Mars.

During the course of the program of flights of landers, there will undoubtedly also be some flyby and orbiter missions which could lead to contamination of Mars by microbes as a result of accidental impact. Because these vehicles would not be sterilized, it may be assumed that every accidental impact would inevitably lead to contamination. If N' is the number of unsterilized spacecraft and I is the probability of accidental impact, then the probability of contamination, $P(C')$, is given by

$$P(C') = N'I \qquad (2)$$

The total probability of contaminating Mars by all spacecraft intended for exploration of the planet, represented by P, is then the sum of $P(C)$ and $P(C')$; that is,

$$P = N\sigma P(g) + N'I \qquad (3)$$

The two quantities of interest are σ and I, but even if the values for all the other quantities were known, it would still not be possible to determine two unknowns from a single equation. Consequently, certain assumptions must be made. Sagan and Coleman suggested, in the first place, that the probability of contamination by sterilized landers, $P(C)$, be taken as equal to that, $P(C')$, for unsterilized orbiters and flybys. In this event, each of these two probabilities is equal to half the total contamination probability:

$$N\sigma P(g) = \tfrac{1}{2}P \qquad (4)$$

and

$$N'I = \tfrac{1}{2}P \qquad (5)$$

There are now two equations, one for σ and one for I, and so a solution is theoretically possible.

The next step in the development is to decide on values for P, $P(g)$, N, and N'. The choice of P, the total permissible contamination probability during the course of the whole program, has been described as a "value judgment." J. O. Light and his associates, in 1966, referred to the selected (maximum) value of P as a "level of commitment." It is "in some sense a measure of the commitment that is deemed appropriate for the avoidance of planetary contamination. . . . The specification [of this quantity] . . . is a difficult and arbitrary task which is not well suited to detailed forms of deductive analysis . . . [and so] it should be formulated upon recommendations stemming from the considered opinion of the scientific community."

The present consensus among scientists is that the "level of commitment" for sterilized landers and unsterilized orbiters and flybys should require a 99.9-percent confidence that the unmanned exploration of Mars will not contaminate the planet with terrestrial organisms. This means that in the whole program the chances of contaminating Mars should be less than one in a thousand; thus, the maximum value arbitrarily assigned to P is 10^{-3}. Whether this is too large or too small no one can say. If it is too small, it includes a factor of safety, although it may hamper to some extent the exploration of Mars by spacecraft. But, taking all the fac-

tors into consideration, the requirement that P shall not exceed 10^{-3} appears to be reasonable.

Next, there is the quantity $P(g)$, the probability that a single terrestrial organism released on Mars will grow and spread over a large area. The value to be assigned to $P(g)$ has been the subject of some controversy, to which reference will be made later. Based on laboratory observations on the degree of survival of terrestrial micro-organisms after freezing and thawing, such as they would undergo each day on the surface of Mars, Sagan and Coleman proposed that $P(g)$ be taken as 10^{-2}. This value will be accepted for the moment.

Finally, the numbers of landers and other spacecraft to be used in the program are required. For purpose of calculation, it may be supposed that 64 landers, that is, $N=64$, would be adequate and reasonable to perform all the desired biological experiments. During the same period, there would be perhaps 25 unsterilized orbiters and flybys ($N'=25$) in the vicinity of Mars. These numbers are estimates based on considerations to be given later. They are somewhat, but not significantly, different from those used by Sagan and Coleman. It is not likely that the suggested values for N and N' will prove to be in error by a factor of more than 2 or so, and this would not seriously affect the final conclusions.

If the values $N=64$, $P(g)=10^{-2}$, and P less than 10^{-3} are inserted into equation (4), it is found that σ should be less than 8×10^{-4}. It would appear, therefore, that the probability of a viable organism reaching the surface of Mars should be kept less than 8×10^{-4} per landing capsule. Similarly, with $N'=25$ and P less than 10^{-3}, equation (5) shows that I, the probability of impact of an orbiting or flyby spacecraft with the surface of Mars, should be less than 2×10^{-5}

The two probabilities thus calculated depend on the postulated value of P, the total probability of contamination of the planet during the whole of the unmanned program, on the estimated value of $P(g)$, the probability that a micro-organism on the Martian surface will grow and spread, on the assumed distribution of the contamination between sterilized landers and other spacecraft which are not sterilized, and finally on the numbers of missions in each of these two categories. Changes in the foregoing quantities will be accompanied by changes in the probabilities σ and I.

According to C. Sagan, who was a member of the Study Group referred to on page 249, the—

> COSPAR recommendations [of May 1964] were based upon an analytic equation for σ . . . obtained from probability theory by Sagan and Coleman. . . . In the deliberations [of the Study Group] . . . each member present . . . gave his best estimate for each numerical parameter entering into the analytic framework. The resulting values of σ were then derived. The discussion iterated in a final value of σ [approximately equal to] 10^{-4}.

It should be noted that σ as defined above is actually the probability per landing capsule that a micro-organism will be deposited on the Martian surface, whereas the COSPAR recommendation is in terms of the probability that a micro-organism will be aboard the spacecraft. Since a micro-organism carried by a landing capsule will not necessarily reach the surface, the latter probability, as expressed in the recommendation, should have been larger than the estimated permissible value of σ. It has been assumed, however, as a factor of safety, that any micro-organism that happens to be on the spacecraft would find its way onto the surface in a relatively short time compared with the duration of the unmanned exploration program.

In a paper entitled "A Discussion of the Planetary Quarantine Constraints," presented

in 1966 by J. O. Light and C. W. Craven of the Jet Propulsion Laboratory, W. Vishniac of the University of Rochester, and L. B. Hall of the National Aeronautics and Space Administration, reference is made to the choice of 10^{-2} for $P(g)$, the probability that an organism reaching the surface of Mars will grow and spread. The authors admit that this value is probably reasonable for random terrestrial micro-organisms released on Mars. They point out, however, organisms remaining on a landing capsule will have been subjected, as will be explained in due course, to a sterilization treatment at a relatively high temperature. Micro-organisms which have survived such heating are less likely to withstand the cold of space and of the Martian surface than are a random group. J. O. Light et al. state:

> These organisms [which reach the Martian surface on a heat-sterilized capsule] should therefore have less chance of surviving, growing, and spreading across the planetary surface to affect biological contamination. It therefore seems appropriate to lower the probability of contamination given the release of an organism on the planet for those viable organisms which have undergone the severe heating inherent in a sterilization cycle. The constant . . . [$P(g)$] should therefore be lowered by an order of magnitude to 10^{-3}, which is still a conservative estimate of the ability of micro-organisms to grow and replicate in the Martian environment.

A decrease in the value of $P(g)$ from 10^{-2} to 10^{-3} would mean that the allowable probability for a landing capsule to carry a single viable organism could be increased correspondingly from 10^{-4} to 10^{-3}. As seen earlier, such a change was implied, although not actually recommended, by the change in $P(g)$ suggested by the COSPAR Consultative Group in 1966.

Inclusion of Other Sources of Contamination

It will be recalled that of three main sources of possible contamination of the Martian surface (fig. 12.1), only the first two were included in the treatment which led to equations (3), (4), and (5). In the discussion referred to above, J. O. Light and his collaborators modified the calculations so as to include the third source, contamination from unsterilized spacecraft resulting from causes other than direct impact. In place of equation (3), they derived the somewhat analogous expression

$$P=NP(r)P(g)+N'[P(r')P(g')+P(r'')P(g'')] \quad (6)$$

where $P(r)$ is the probability that an organism will be released from a space vehicle and $P(g)$ is the probability that such an organism will subsequently grow and spread on the surface of Mars. The quantities N, $P(r)$, and $P(g)$ refer to sterilized landers, N', $P(r')$, and $P(g')$ to the impact of an unsterilized vehicle, and $P(r'')$ and $P(g'')$ to organisms which reach the surface from such a vehicle by ways other than direct impact. The probabilities $P(r)$ and $P(r')$ are essentially equivalent to σ and I, respectively, of the simpler treatment given earlier.

For purposes of calculation, it was postulated that manned exploration of Mars would supersede unmanned missions around the year 2000. In the meantime, there would be 16 opportunities, coinciding with the 16 oppositions, for sending unmanned spacecraft to Mars. If the United States and Russia each successfully launched two landing vehicles at all 16 opportunities, the number of landers (N) would be 64. This is the number used in the earlier calculations. Furthermore, it was assumed that of these flights, 25 would include an unsterilized orbiter or flyby vehicle, so that N' is 25.

The value of $P(g)$ may be taken as 10^{-3} in accordance with the arguments already

given, but for $P(g'')$, which applies to organisms on an unsterilized vehicle, the value of 10^{-2}, originally used for $P(g)$, may be retained. In the earlier calculations, it was accepted that $P(g')$, the probability of contamination resulting from an accidental impact, was unity. Because of the generally hostile nature of the Martian surface, however, it is suggested that the value of $P(g')$ be reduced [2] to 0.5, to "take into account the possibility that there . . . may be large areas on Mars where terrestrial organisms will find it impossible to survive and grow." Thus, equation (6) can be written as

$$P=64[P(r)\times 10^{-3}]+25[(P(r')\times 0.5)+(P(r'')\times 10^{-2})] \quad (7)$$

Given a certain "level of commitment," for example, that P shall be kept less than 10^{-3}, this equation sets maximum values on the permissible release probabilities, $P(r)$, $P(r')$, and $P(r'')$. Each of these may be obtained by setting the other two equal to zero. It is found in this way that the possible ranges of values are

$P(r)$ from 0 to 1.56×10^{-2}
$P(r')$ from 0 to 8×10^{-5}
$P(r'')$ from 0 to 4×10^{-3}

to satisfy the requirement that P should not exceed 10^{-3}. Within these limits for $P(r)$, $P(r')$, and $P(r'')$, however, there are an infinite number of possible values for the three probabilities that will all lead to P being less than or equal to 10^{-3}.

The only practical way out of this situation is to assume reasonable values for any two of the three $P(r)$'s and then to calculate the third by means of equation (7) with P taken to be less than 10^{-3}. Thus, if the implied COSPAR values are assumed, $P(r)$, the probability that a viable organism will be released from a sterilized lander, may be set at 10^{-3}, and $P(r')$, the equivalent probability in the accidental impact of an unsterilized vehicle, is 3×10^{-5}. It is then found from equation (7) that $P(r'')$ should be less than 2.3×10^{-3}. The probability of contamination of the surface of Mars by unsterilized vehicles because of reasons other than accidental impact should thus be kept below 2.3×10^{-3} to satisfy the postulated level of commitment. It should be pointed out that this value of $P(r'')$ is dependent, among other things, on the values chosen for the probabilities $P(r)$ and $P(r')$ which are themselves subject to readjustment. If the value of $P(g')$ is 1.0, instead of 0.5, then $P(r'')$ would have to be kept below 0.76×10^{-3}.

Arguments for Relaxation of Constraints

During 1967, some arguments were presented in favor of a substantial relaxation of the quarantine constraints recommended by COSPAR. Although they have received little support, it is, nevertheless, of interest to review the situation. In an area in which subjective judgments play such an important role, changes of opinion are not uncommon. As already seen, the COSPAR constraints appear to assume, at least tacitly, that essentially every microbe remaining on a sterilized landing capsule will be released on the Martian surface.

N. H. Horowitz and his associates contend, however, that this would not be so. The commitment to a total probability (P) of less than 10^{-3} for the contamination of Mars during the period of unmanned exploration could consequently be realized with the probability of a single viable organism being aboard a lander appreciably larger than the 10^{-3} implied by the 1966 COSPAR recommendation.

[2] The COSPAR Consultative Group did not concur in this change and suggested that $P(g')$ be retained as unity.

If all the exterior surfaces of a landing capsule and its components are sterilized, as can be done relatively simply (p. 259), the only organisms remaining are those trapped in the interiors of solid parts, such as transistors, capacitors, resistors, cores, and so on. In any event, as a result of manufacturing and testing procedures, the degree of internal contamination of many such parts is very small, and for others it could be kept low by taking simple precautions.

If the capsule lands without damage on the surface of Mars, the trapped micro-organisms, if any, can be released only by diffusion through the solid to the surface or as a result of the solid being eroded away by wind-driven particles. It is argued that neither of these factors would be significant. Suppose, on the other hand, that the capsule intended for landing crashes on the surface of Mars and is disintegrated. "Given the low degree of contamination of electronic parts," state Horowitz et al., ". . . it can be shown that extensive fragmentation of the lander into submillimeter-sized particles must occur before there is a significant probability of releasing micro-organisms onto the surface." The conclusion drawn is that, regardless of whether the capsule lands intact or crashes, the probability of the Martian surface becoming contaminated is very small.

Furthermore, even if a terrestrial micro-organism reached the surface of the planet from the unsterilized interiors of the solid components of a landing capsule, the chances of spreading so as to contaminate Mars are believed to be extremely small. "The Martian environment is an extremely hostile one by terrestrial standards. The dryness, lack of oxygen, mean low temperature, and high ultraviolet flux make it an inhospitable place for any migrant micro-organism from Earth." This does not mean that indigenous life, adapted to the conditions on Mars, cannot thrive. But the probability that a terrestrial organism will spread over the planet is believed to be much less than the 10^{-2} used by C. Sagan and S. Coleman in their calculations. It would presumably also be regarded as less than the 10^{-3} probability (for $P(g)$) adopted by J. O. Light and associates in 1966.

In summary, Horowitz and his collaborators state:

> A distinction needs to be made between micro-organisms trapped in solids and those on exposed surfaces of landed spacecraft. Surface sterility is an unconditional requirement . . . . Sterilization of the interior of solids to the extreme level recommended by COSPAR, however, is based on the assumption that entrapped organisms constitute a substantial hazard to the ecology of Mars. This assumption now seems unjustified, and the need for a high degree of interior sterility is doubtful.

In a rebuttal of the foregoing arguments, C. Sagan, E. C. Levinthal, and J. Lederberg point out that there is not enough known about the possibility of wind erosion of solids on Mars or of the degree of fragmentation of a spacecraft accompanying a high-velocity impact to justify any changes in the quarantine requirements. The effects of erosion, for example, could be substantial, and the crash of a landing capsule on Mars could conceivably lead to complete pulverization. As regards survival and growth of terrestrial organisms, attachment to soil particles could provide protection from ultraviolet radiation. Moreover, direct landing in or transport by the wind to a favorable microenvironment would be conducive to growth. Another possibility is that, in the event of a crash, the nutrient solutions, carried by the lander to test for life on Mars, would provide an excellent medium for the establishment of a colony of terrestrial organisms. Some of these might become adapted to the Martian conditions and would spread over the surface.

On the whole, Sagan and his associates conclude:

> The . . . 10^{-2} . . . probability that a terrestrial micro-organism deposited on the surface of Mars will grow and contaminate the planet may in fact be too low rather than too high. . . . We anticipate that . . . over a long time scale we will land not just at an average location on the planet, but at a place where terrestrial [and presumably Martian] micro-organisms are most likely to grow. Our subsequent concern is not necessarily with growth over large areas of the planet, but rather with growth at those favorable sites chosen for subsequent investigations. . . . We believe that the probabilities of spacecraft erosion and fracture, survival [of micro-organisms] from the ultraviolet flux, and subsequent ultimate deposition in a warm and wet locale on Mars may be rather high. If sterilization standards are to be relaxed, we must be quite certain—much more certain than we can be in our present [1967] state of ignorance—that such will not be the case.

An entirely different reason for easing the quarantine requirements for spacecraft to be used in the exploration of Mars has been given by B. C. Murray and his colleagues. It is claimed that the U.S.S.R. spacecraft Zond 2 which, according to Russian statements, was intended as a Martian flyby, actually crashed on the surface of the planet in August 1965. Because Zond 2 was not sterilized, the surface of Mars is already contaminated and there is consequently no point in setting stringent sterilization requirements for subsequent missions.

The only available data of the trajectory of Zond 2 were those obtained with the large (250-foot) radio telescope at the Jodrell Bank Observatory in England. Bernard Lovell, the director of the observatory, reported that Zond 2 was located on December 2, 1964, the day after it was launched. Its path was tracked on successive days during most of January 1965, and the tracking was continued, although at longer intervals, through February and March, and the first week of April. The increased time intervals were not the result of weakness in the radio signals, but because the transmissions from the spacecraft occurred less frequently. The last contact made by Jodrell Bank with Zond 2 was on April 7, 1965. A search for signals was made around the date of predicted closest approach to Mars on August 6, 1965, but none was received. If there had been any radio signals from the spacecraft, they would have been detected by the Jodrell Bank telescope, so it must be assumed that the transmitter on Zond 2 failed on or soon after April 7, 1965.

The trajectory calculated from the tracking data obtained at Jodrell Bank prior to this date indicated that Zond 2 was on an apparent collision course with Mars. Upon being informed of this fact, M. V. Keldyesh of the U.S.S.R. replied that the spacecraft would "pass within 1500 kilometers of the planet." In reviewing the matter, Lovell stated: "Clearly the Soviet scientists were in a position to make far more precise calculations about the orbit without some of the assumptions that had to be made in the Jodrell calculations. Further, it must be emphasized that a miss distance of 1500 kilometers was within the accuracy of our . . . calculations."

It is by no means certain, therefore, that Zond 2 did impact on Mars. If Russian scientists have any information indicating that such impact did occur, it is very improbable that they will release it. Consequently, the situation will remain unresolved. In any case, even if a limited area on Mars is already contaminated, this is not felt to be a justification for contaminating other parts of Mars. The majority opinion of scientists in the United States is probably that quarantine standards should not be lowered on the basis of what is known about Zond 2.

Model for Planetary Quarantine

It is evident from the foregoing discussion that there are many uncertainties in-

volved in the establishment of the requirements for the sterilization of Martian landing capsules and for the avoidance of accidental contamination by unsterilized spacecraft and their components or ejecta. In the hope of reaching more satisfactory conclusions, a proposal has been made to COSPAR for a detailed mathematical model of planetary quarantine requirements. By taking into account all conceivable factors that may contribute to the contamination of Mars, it should be possible to make better estimates of the various requirements than those based on the simpler considerations described earlier in this chapter.

The total probability, P, of contaminating Mars during the entire period of unmanned exploration may be represented by the sum of three terms:

$$P = P(C) + P(C') + P(C'')$$

where $P(C)$ is the probability of contaminating the planet during the prescribed period by spacecraft which have been sterilized, $P(C')$ is the corresponding probability for impacting unsterilized spacecraft, and $P(C'')$ is the probability of contamination by an unsterilized vehicle from all sources other than accidental impact of the vehicle itself.

Each of these three probabilities is further subdivided in terms of (1) the probability of survival of an organism on a spacecraft after its journey through space, from Earth to Mars; (2) the probability of release of a viable organism on the surface of the planet after it has survived the journey through space; and (3) the probability that an organism which survives the journey through space and is released on the planet will subsequently grow and spread. In evaluating the terms $P(C')$ and $P(C'')$, an additional probability is included to allow for guidance failure, leading to accidental impact on the planet.

If the numbers of sterilized landers and of unsterilized space vehicles are known, estimates can be made from $P(C)$, $P(C')$, and $P(C'')$, along the lines indicated earlier in this chapter, of the probabilities, $P(C_i)$, $P(C_j')$, and $P(C_j'')$, that an individual lander i or an individual unsterilized vehicle j will contaminate the planet.

The recommendation has been made that—

> COSPAR be responsible only for specifying P and finding out the anticipated number of missions which each nation is expected to launch [to Mars] . . . during the period of unmanned exploration. Each nation given [this] information . . . should then determine . . . $P(C_i')$, $P(C_j')$, and $P(C_j'')$ for its flights by a conservative approximation [based on estimates of the various probability subdivisions described above].

IMPLEMENTATION OF STERILIZATION REQUIREMENTS

Decontamination and Sterilization

The sterilization requirements for a Martian landing capsule, weighing several hundred pounds and including many thousand individual parts—electronic, electrical, and mechanical—are more severe than those commonly encountered in medical practice. A surgeon's scalpel, for example, must be sterile on the outside, but it is immaterial whether or not its interior is sterile. The chances of the scalpel disintegrating during an operation are extremely remote. A space vehicle intended for landing on Mars, however, must be completely sterile, at least to the extent of a probability of less than one in a thousand that there will be even a single viable organism anywhere on or within its many component parts. Considerable effort has been devoted, especially in the United States, to the problem of spacecraft sterilization.

The fabrication of a sterile spacecraft in-

volves both biological decontamination, to bring about a substantial decrease in the number of microbes and spores, and sterilization, to kill all the organisms present. The procedures employed (or proposed) for the decontamination and sterilization of space vehicles and their components are based on the use of the following: (1) chemical sterilants—gaseous, liquid, or solid; (2) electromagnetic radiation—ultraviolet or X-rays; (3) heat. In addition, there is often a spontaneous decay of micro-organisms if stored under sterile conditions; this may be significant in some circumstances.

Chemical Sterilization by Gas

The most common gas used in the sterilization of space-vehicle components is ethylene oxide, $(CH_2)_2O$, often abbreviated ETO. It is gaseous at room temperature, but at atmospheric pressure it liquefies when cooled to 10.8° C. With a relatively small increase of pressure, ethylene oxide can be liquefied at ordinary room temperature. The gas appears to be effective in killing micro-organisms, including spores, of all kinds. Nevertheless, it is only moderately toxic to human beings; concentrations of 5 to 10 parts of ethylene oxide per million of air are reported to be harmless for breathing over several years. Furthermore, it has essentially no deleterious effects on metals, most plastics (or polymeric substances), and other materials used in the fabrication of space vehicles and their component parts.

In the pure state, ethylene oxide is readily flammable and the gas can form an explosive mixture with air. Consequently, when employed as a sterilizing agent a nonflammable diluent is always added. Carbon dioxide has been used for this purpose in a mixture of 10 percent by weight of ethylene oxide and 90 percent of carbon dioxide. But containment of the mixture in liquid form, under a substantial pressure, requires the use of heavy steel cylinders.

In the United States, the diluent is generally Freon-12 (dichlorodifluoromethane, CCl_2F_2) which liquefies at −29.8° C at ordinary pressure. The sterilizing mixture available commercially consists of 12 percent by weight of ethylene oxide and 88 percent of Freon-12. Although the Freon costs more than carbon dioxide, the containers are inexpensive and easy to handle. The ethylene oxide-Freon mixture is maintained as a liquid under moderate pressure at ordinary temperature. Upon opening a valve on the containing cylinder, the gas emerges. An advantage of this mixture over the one with carbon dioxide is that the Freon diluent also has germicidal properties.

Russian scientists have reported that a mixture of 40 percent by weight of ethylene oxide with 60 percent of methyl bromide (CH_3Br) is nonflammable and is particularly effective as a sterilizing agent. Not only does this contain more ethylene oxide than the common mixture with Freon but methyl bromide is itself a good sterilant. Furthermore, there is some indication that, in certain proportions, ethylene oxide and methyl bromide may exert a mutual synergistic effect; the sterilizing action of the mixture is then greater than the sum of the contributions of the separate components. Because of the larger proportion of ethylene oxide and the presence of methyl bromide, the mixture sterilizes much more rapidly than does the ethylene oxide-Freon mixture when used at the same atmospheric pressure.

A certain amount of water vapor should be present for ethylene oxide to be effective as a sterilizing material. Bacterial spores which have been partly desiccated acquire a degree of resistance to sterilization until they reabsorb sufficient moisture. Ethylene oxide

mixtures are generally employed at 30 to 50 percent relative humidity, meaning the water vapor content is between 30 and 50 percent of the amount required to produce saturation (condensation of liquid water) at the given temperature. Common concentrations of ethylene oxide for sterilization purposes are 300 to 500 mg/l of air, and temperatures may range from about 25° to 40° C or more.

Increase of ethylene oxide concentration and of temperature are each able to increase the rate of sterilization. Roughly speaking, a rise of 10°C in the temperature decreases the sterilization time to somewhat less than half. Thus, bacteria or spores which require treatment with ethylene oxide for 10 hours at 25°C can be sterilized in less than 4 hours at 35°C. A longer processing time may be advantageous, however, when a complex assembly is being sterilized and the gas has to penetrate into many narrow spaces.

One of the outstanding aspects of ethylene oxide as a sterilant is its good penetrating power. Because of this property, components to be sterilized can be wrapped in paper and they can be handled subsequently without risk of recontamination. Although ethylene oxide can readily penetrate into cracks and crevices, it cannot get much below the surface of nonporous solids. Its main use is therefore as a surface sterilant. In most cases, the residual ethylene oxide on the sterilized object or material dissipates quite rapidly. One important exception, however, is rubber. About a day must be allowed for removal of ethylene oxide from rubber gloves, otherwise the wearer may suffer blistering of the skin.

Liquid Chemical Sterilants

Liquids can serve the same purpose as gases for sterilizing the surfaces, but not the interiors, of solid components. If employed under proper conditions, liquids act more rapidly than gases, mainly because of the higher concentration of the sterilant. As a result of the fairly large surface tension of water, aqueous solutions do not wet surfaces readily, especially if they have a thin film of oil or grease, as is generally the case. Consequently, organic liquids, with small surface tensions, are usually employed as solvents. A possible alternative is to add a detergent to the aqueous solution to improve its ability to wet surfaces. It should be noted that several so-called liquid sterilizing agents are not active against bacterial spores, but they can, nevertheless, serve to reduce the total load of microorganisms.

Cleaning, brushing, or immersion in a liquid sterilizing solution has been suggested for the decontamination of small electronic parts prior to assembly. Long immersion, to produce sterilization, however, has resulted in deterioration of the components in some instances. One of the best liquid sterilants has been found to be a solution of formaldehyde in methyl alcohol, with or without the addition of water. Brushing with a liquid fluorocarbon containing phenol and swabbing with ethyl alcohol have been employed to decontaminate the components of electronic subassemblies.

In the United States, liquid sterilants have found relatively little application, except for the decontamination of small parts. V. I. Vashkov and his collaborators in the U.S.S.R., however, reported in 1967 that they had achieved effective destruction of both bacteria and spores by immersion of components for a sufficient time in a dilute aqueous solution of hydrogen peroxide to which a detergent had been added to decrease the surface tension. Wiping the surfaces of solids with such a solution was also used for decontamination purposes.

Although solid materials are not used for direct sterilization, they can exert the same

function in an indirect manner. One of the major problems in the sterilization of space vehicles is the destruction of organisms in the interior of solid components. As seen above, neither gaseous nor liquid sterilants are effective in this respect. The problem could be solved, or at least made less severe, if a material with sterilizing properties could be incorporated into the solids from which the component is fabricated.

Solid Chemical Sterilants

Paraformaldehyde, a solid polymer of formaldehyde and a good sterilant, can be incorporated into plastics, electrical potting compounds, and solid rocket fuels. Some plastics, the epoxy resins, for example, are themselves toxic to both bacterial cells and spores. In other cases, the compound, such as antimony pentachloride, added as a catalyst in the production of the plastic (polymeric) material has germicidal properties. Soviet scientists have expressed a special interest in the development of "materials exhibiting . . . high . . . activity for the most complete destruction of microflora on the surface and in the interior of parts of space research apparatuses. It is necessary," they say, "to encourage new and further explorations in this field."

Sterilization by Radiation

It has been mentioned earlier that the great majority of micro-organisms are unable to survive exposure to ultraviolet rays. Such radiations with wavelengths around 2600 Å are especially lethal because they are absorbed by and decompose the nucleic acids which are essential to life. Most of the ultraviolet rays from the common mercury-vapor lamp have wavelengths in the region of the mercury line at 2537 Å. Consequently, mercury vapor lamps have found some use in industry for the destruction of both bacteria and fungi.

Unfortunately, as noted in chapter X, micro-organisms are often protected from ultraviolet light by small particles to which they are attached or even by surface roughness. Complete sterilization, even of a surface, is thus not always possible. Nevertheless, because ultraviolet radiation can usually destroy exposed micro-organisms very quickly, it might be utilized in decontaminating procedures to reduce the microbial load of components of a space vehicle. Such decontamination, as will be seen later, would ease to some extent the problem of sterilizing the complete vehicle after assembly.

At best, ultraviolet rays will only kill micro-organisms on the surface of a solid, because in cases of interest they are unable to penetrate into the interior. The related electromagnetic radiations of shorter wavelength, namely, X-rays and gamma rays, have both considerable penetrating power and sterilizing ability. These two radiations, sometimes referred to as "ionizing radiations," differ essentially only in their mode of production. X-rays are obtained from conventional X-ray machines by bombarding a target of a heavy element, such as tungsten, with a beam of high-energy electrons, whereas gamma rays are emitted spontaneously by many radioactive substances. Cobalt sources of gamma rays containing radioactive cobalt-60 are readily available. Because of their ability to penetrate some distance into solids, X-rays and gamma rays can destroy organisms in the interior of solid components of space vehicles.

High-energy (short-wavelength) ionizing radiations are being used for the sterilization of various hospital supplies and also for the preservation of meat and other foodstuffs. The radiation doses required to kill all micro-organisms are, however, very large, usually from 2.5 to 5 million rads (p. 215). Of the available sterilizing agents, only two—high-energy

radiation and heat—are capable of killing micro-organisms on the surfaces of spacecraft components and also in the interiors. As will be seen later, the decision has been made in the United States to employ heat sterilization for the components and complete assembly of Martian landing capsules. There is no point, therefore, in using radiation as the sterilizing agent at any stage. Radiation could, of course, be used as an alternative to heat for both intermediate and final sterilization, but, according to L. B. Hall, Planetary Quarantine Officer of the National Aeronautics and Space Administration, "radiation is expensive, hazardous, complex, and damages many materials more than does heat."

The approach to sterilization of planetary landing vehicles taken in the U.S.S.R. is different from that in the United States. The final sterilization of the assembled space vehicle is to be performed by means of ethylene oxide gas. Because this will destroy only those micro-organisms on the surface, it is proposed to make sure that all components are sterile internally before the final assembly. As many components as possible will be sterilized by heat, but for those which are damaged by high temperatures, it is proposed to make use of radiation.

In 1966, A. K. Afstafyeva and her associates in the U.S.S.R. stated:

> It has been decided to sterilize electroradio [electronic] apparatuses with the aid of . . . gamma rays. . . . Diodes, condensers, resistors, relays, [and other components] . . . which meet technical requirements may be subjected to sterilization in the course of exposure with a dose of 2.5 [million rads] . . . of gamma rays. [This dose] . . . caused no essential changes of electrical parameters within the restricted limits specified in the technical instructions provided by the manufacturers.

Experience in the United States, however, does not appear to substantiate this conclusion.

Sterilization by Heat

Heat is the most effective and reliable method of killing microbes and spores, both on the surface and in the interior of solid components of a spacecraft. Moist heat has been used for many years for sterilization in hospitals, in bacteriological laboratories, and in the food industry, but relatively little has been done in connection with dry-heat sterilization, which is less effective. Furthermore, such studies as had been made with dry heat mostly involved temperatures above 150° C, but many spacecraft components would be unable to withstand such high temperatures. Around 1961, when the National Aeronautics and Space Administration decided that capsules for landing on Mars would be sterilized by heating in a dry atmosphere, systematic investigations were started to determine the effect on micro-organisms of temperatures in the range of about 110° to 160° C.

Dry-heat sterilization is evidently a fairly complex phenomenon. Some simple, but approximate, principles will be described first, and necessary modifications will be considered later. Before a final decision can be made concerning the temperature and duration of the heat sterilization cycles to be used for Martian landing capsules and their components, considerably more information will be required. Efforts are being made to obtain this information.

It is well known to bacteriologists that not all bacteria or bacterial spores of the same kind in a colony respond equally to a given sterilizing agent. Some are killed very quickly, whereas others are more resistant and survive for a longer time. This behavior is quite general and has been observed in connection with sterilization by heat as well as by other agents. To avoid damage by excessive heating, the manner in which the bacterial load changes with time is especially important.

The results given below are applicable to both bacterial cells and spores, but major attention has been given to spores because they are more difficult to destroy by heat, as well as by other means.

If a sample of bacterial spores of a particular species is maintained at a constant temperature, say about 130° C, in a dry atmosphere, commonly nitrogen gas, the number of viable spores present decreases in an approximately logarithmic (or exponential) manner. In other words, if the number of spores surviving after a given time of heating at constant temperature is plotted on a logarithmic scale against the time (on a linear scale), the points will fall roughly on a straight line. The full line in figure 12.2, shows how the spore population decreases from 10^8 (log $10^8=8$) at the beginning to 1 (log $1=0$) after 24 hours. The significance of the dotted extension of the line will be explained shortly.

One of the consequences, which is evident from figure 12.2, of the logarithmic decay in the number of spores with time during the sterilization process, is that it takes the same time to decrease from 10^8 to 10^7, as it does from 10^7 to 10^6, from 10^6 to 10^5, and so on.

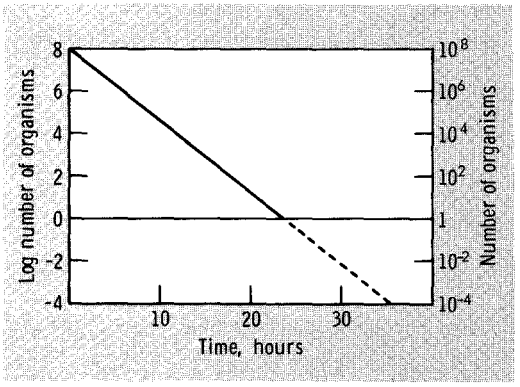


FIGURE 12.2. Change in number of micro-organisms with time in dry heating at constant temperature.

In other words, no matter how many organisms are present, it always requires approximately the same time of heating at a given temperature to decrease the number of spores, or micro-organisms in general, to one-tenth of the original quantity. Incidentally, this type of behavior is also characteristic of radioactive decay and of first-order chemical reactions.

The time required for a population of bacterial cells or spores to decrease to one-tenth is called the decimal reduction time, and is represented by the symbol D. A decrease to one-tenth in actual number means a decrease of one unit in the logarithm of the number. Hence, D may be defined as the heating time at a specified temperature in which the logarithm of the number of survivors decreases by unity. If N_0 is the number of micro-organisms present before heating and N_t is the number surviving after time t, the change in the logarithm is $\log N_0 - \log N_t$, which is the same as $\log (N_0/N_t)$. To a first approximation, therefore, it follows that

$$t = D\left(\log \frac{N_0}{N_t}\right) \qquad (8)$$

Thus, if the decimal reduction time D is known for a given organism, at the particular temperature at which dry-heat sterilization is performed, it is possible to calculate the time t required to decrease the number of organisms from the initial value N_0 to any specified final value N_t. (Some experimenters include the term -1 in the parentheses of equation (8) to allow for the almost instant death of the most sensitive organisms. This adjustment will not be used, however, in subsequent calculations.)

After the number of organisms decreases to one, that is, $N_t=1$ and $\log N_t=0$, no further observations can be made. But the

dotted line in figure 12.2 which is an extrapolation of the experimental data, is taken to represent the probability of survival of a single organism. On this basis, equation (8) can be modified to obtain an indication of the number of decimal reduction times required to decrease a certain number of spores N_0 to such an extent that the probability of one surviving is P_s. Thus, replacement of N_t in equation (8) by P_s gives

$$t=D\left(\log \frac{N_0}{P_s}\right) \qquad (9)$$

Suppose a landing capsule, which is estimated to contain a total of 10^5 organisms, is to be sterilized by dry heating at a particular temperature until the probability of a single organism remaining is 10^{-3}. The heating period in terms of decimal reduction times would be log (N_0/P_s); that is, log $(10^5/10^{-3})$ $=\log 10^8=8$. Thus, about eight decimal reduction periods, for the most resistant organisms present, at the given temperature would be required to achieve the objective of the sterilization process.

The decimal reduction time, as already implied, depends on the temperature at which the dry-heat sterilization is being performed. The time decreases with increasing temperature and, as a general rule, the plot of the logarithm of the decimal reduction time against the temperature is approximately a straight line (fig. 12.3). The slope of the line varies to some extent with the nature of the micro-organisms, but, very roughly, it requires an increase of about 25° C in temperature to decrease the decimal reduction time to one-tenth its original value. This result may be stated in another way: The decimal reduction time is decreased to between one-half and one-third by every 10° C increase in temperature.

For example, if the decimal reduction time is 25 hours at 120° C, it would be something like 10 hours at 130° C, and roughly 4 hours at 140° C. These figures must not be taken as being exact, but they are quoted to provide a general indication of the effect of temperature on the length of the sterilization cycle.

The variation of the decimal reduction time with temperature, which must be determined by experiment for the bacterial spores likely to remain in a space vehicle, is of importance from the practical sterilization standpoint. Heating at high temperature even for a short time often decreases the functional reliability of spacecraft components, but heating at a lower temperature for a much longer time may cause no damage. A given component may become unreliable after heating for 4 hours at 140° C, but its operation may be unaffected when heated at 120° C for 24 hours or longer. Both sets of conditions would be capable, however, of causing the same degree of sterilization. It may prove necessary, therefore, to compromise between time and temperature in deciding on the conditions to be used in the sterilization of the components and the complete instrumented

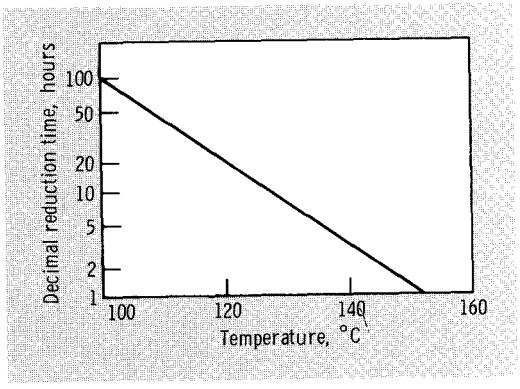


FIGURE 12.3. Effect of temperature on decimal reduction time in dry heating.

capsule that is intended for landing on Mars.

The general principles of dry-heat sterilization outlined above are substantially correct when the number of micro-organisms present is large or moderately large. In these circumstances, the decrease in the bacterial or spore population with time of heating is close to logarithmic, as implied by the straight line in figure 12.2. But as the number of survivors becomes smaller, there is often a deviation from the linear plot. This deviation, shown in figure 12.4, is such as to make the number of organisms surviving larger than expected. The time to produce a required degree of sterilization estimated from equation (9), that is, from the dotted line in figure 12.2, would then prove to be too short.

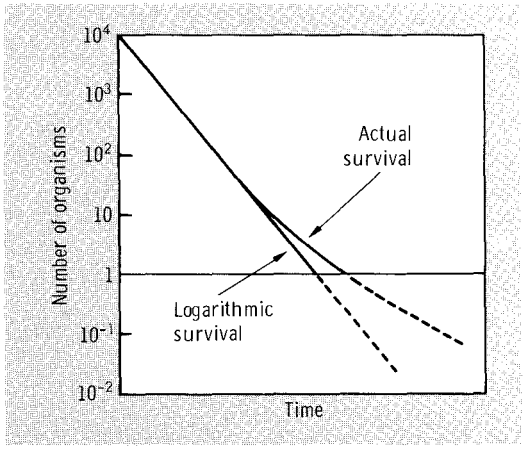


FIGURE 12.4. Deviation from ideal logarithmic survival behavior in sterilization by dry heating.

One way of overcoming this difficulty would be to extend the heating time, somewhat arbitrarily, to include a factor of safety. Apart from the uncertainty involved, such a procedure is undesirable because it might lead to unnecessarily prolonged heating with accompanying damage to the components of the space vehicle. A more scientific approach is to try to get a better mathematical representation of the survival of micro-organisms in heat sterilization.

In 1967, S. Schalkowsky and R. Wiederkehr in the United States proposed a model for the survival of micro-organisms based on the postulate that the prior history of heating affects the future resistance of the organisms to heat. It is supposed that there is a cumulative heating time effect; the longer the organism has been exposed to heat, the greater is the probability of its surviving further. On the basis of this assumption, it is found that the natural logarithm (to the base e) of the survival times of individual micro-organisms selected at random from a given population should follow what is called a "normal" or Gaussian distribution.

A fairly complicated mathematical expression, with two adjustable quantities to represent the resistance of the organisms to heat, can then be derived relating the initial population (N_0) to that at any later time (N_t). This expression is a representation of the "two-parameter log-normal" model of the survival of micro-organisms when heated. One of the parameters, that is, one of the adjustable quantities, depends mainly on the type of organism, whereas the other is determined by the heating temperature.

The two-parameter log-normal model has been found to describe in a very satisfactory manner the survival of micro-organisms, especially after relatively long heating times when the number remaining is small. This is just the condition in which the simple logarithmic survival model fails (fig. 12.4). The log-normal model can be improved still further by introducing a third parameter, leading to the "three-parameter log-normal" model. The third adjustable quantity, determined by the environment in which the organisms are heated, eliminates deviations from the two-parameter model during the initial heating period.

It has been known since 1962, from the work of C. W. Bruch and his coworkers in the United States, that the survival of bacterial spores during heat sterilization is dependent on the environment. Thus, using the decimal reduction time as the criterion, it was found that spores embedded in a solid rocket propellant were more resistant to heat than the same spores deposited on sand. Furthermore, the spores on sand were more resistant than those on filter paper. From other studies, it has been found that the nature of the atmosphere, whether air, helium, or a vacuum, also determines the survival of micro-organisms when heated. Another aspect of the environment is the quantity of moisture present in the organisms. Thus, the survival of spores enclosed in the plastic Lucite was found to vary with the proportion of water.

Obviously, sterilization by dry heat is more complex than it was once thought to be. The three-parameter log-normal model, which represents the survival of micro-organisms better than does the simple logarithmic equation, appears to be an advance because it identifies factors dependent on the organism, the temperature, and the environment. Much more experimental work needs to be done to implement the model, but its use, in the words of Schalkowsky and Wiederkehr, "would lead to a more rational basis for selecting safety factors for sterility assurance and for relating these . . . factors to the explicit conditions on which [sterilization] requirements are to be based."

Natural Dieaway of Micro-Organisms

When placed in a sterile environment at ordinary temperature, many, if not all, organisms that do not form spores decrease in number at a moderately rapid rate. This phenomenon is called dieaway, or dieoff. When a bacteriologically clean surface is exposed to the atmosphere, it soon becomes contaminated as a result of what is known as bacterial fallout. But the amount of contamination does not increase indefinitely. It reaches a maximum value in a few days and then remains almost constant. One factor, at least, contributing to this behavior is that the bacteria are dying away about as fast as additional ones are being deposited by fallout from the atmosphere.

Bacterial spores and spore-forming bacteria die away very slowly under normal conditions. There are indications, however, that in a very high vacuum, similar to that existing in space, the dieoff of unprotected spores is quite significant. If this proves to be the case, advantage may be taken of the dieaway during the journey of a space vehicle to Mars.

Decreasing Contamination in Fabrication

The policy adopted by the National Aeronautics and Space Administration in the United States is that, after assembly, instrumented capsules intended for landing on Mars should be subjected to a final sterilization by dry heating. The temperature and time of heating will be chosen so as to reduce the probability of a single viable organism remaining aboard the capsule to less than 10^{-3}, without impairing the ability of the components to fulfill their intended purpose. As seen above, the duration of the heating period, at a given temperature, necessary to achieve the required degree of sterilization depends on the bacterial load prior to heating. Consequently, in order to reduce the heating time, and thereby ensure greater operational reliability of the instruments and other equipment, it is desirable to maintain a low level of contamination by micro-organisms during the various stages of fabrication and assembly of the components.

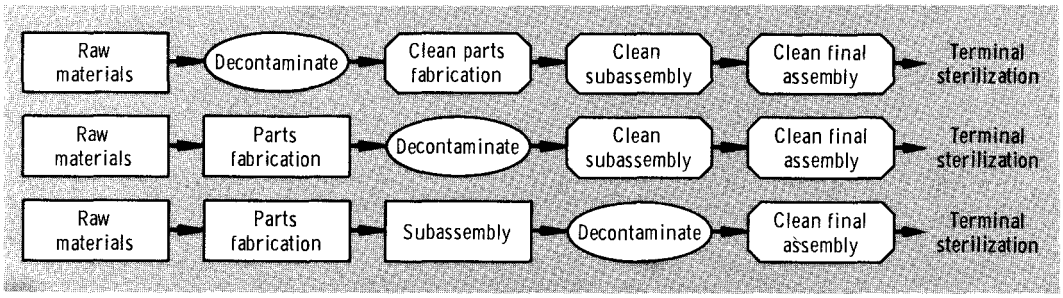


FIGURE 12.5. Different ways of minimizing bacterial load prior to dry heat sterilization.

Three general ways of minimizing the bacterial load of the capsule before the terminal dry-heat sterilization are indicated in figure 12.5. In the first, the materials used in fabricating all parts would be sterilized, or at least decontaminated biologically. Then all stages of assembly would be carried out under specially "clean" conditions designed to keep subsequent contamination to an absolute minimum. Maintenance of the necessary levels of biological cleanliness presents many problems, and operations are inevitably much slower than in normal circumstances. The procedure of clean handling throughout is regarded as being too costly and time consuming for serious consideration.

The second possibility is to manufacture the individual parts of the capsule under ordinary industrial conditions. The parts are then sterilized or decontaminated, and all subsequent stages—combination into subsystems and finally into the complete assembly—are performed in a biologically clean environment. This method is much less complex than the one described above, because fabrication of the many thousands of parts that go into a capsule would not require special facilities. Nevertheless, production of the subsystems and their assembly into the capsule would still be expensive.

A simpler and more practical approach is the third; this is the one which is currently under study in the United States. Parts and subsystems are fabricated under the moderately clean conditions generally used in the production of hardware for space vehicles. This would include reducing the amount of dust in the atmosphere and keeping all surfaces reasonably clean. Materials which are toxic to micro-organisms might be included in some solid components, as indicated earlier. After decontamination, the subsystems would be assembled into the complete system in a biologically clean facility.

There is, in addition, an intermediate step not shown in figure 12.5. As part of the quality-assurance requirement, all subsystems would be subjected to the action of ethylene oxide gas and then to dry heating before undergoing operational tests. The heating time and temperature would be sufficient to result in internal sterilization of all solid components and this condition would be maintained throughout all later stages.

A proposed plan for implementing the foregoing procedure is shown in figure 12.6. The subsystems, produced under reasonably normal clean working conditions, are treated with an ethylene oxide-Freon mixture, of about 40 percent humidity, for a time sufficient to destroy all surface organisms. Next, the subsystems are heated in dry nitrogen in three

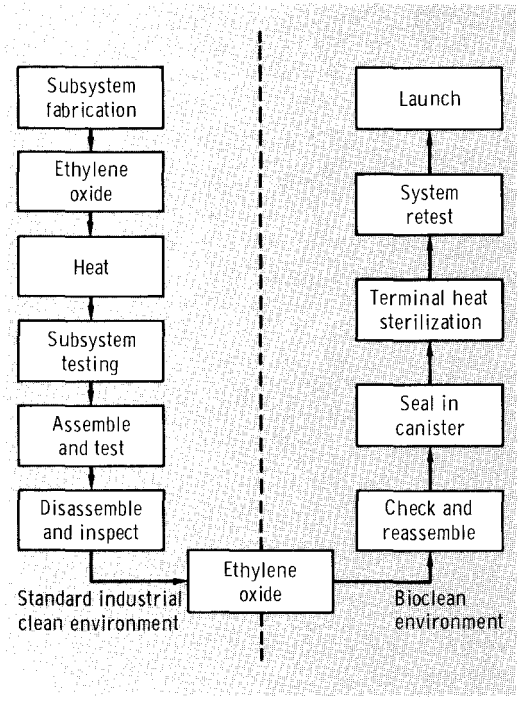


FIGURE 12.6. Proposed scheme for sterilization of spacecraft.

or more cycles at temperatures up to 145°C and for a total time that may extend for several days. Although the primary purpose of these treatments is to make sure that the operational effectiveness of the components is not adversely affected by moist ethylene oxide and by high temperature, they result in complete sterilization. The subsystems are now tested, and if found satisfactory, they are assembled into the complete system which is also tested. The system is next disassembled and inspected. As part of the testing program, samples of solid components could be ground into a fine powder and examined for the presence of micro-organisms to make sure they are sterile.

So far, all operations at the left of figure 12.6 have been performed in a normal, moderately clean environment. The subassemblies are now transferred, by way of a treatment with ethylene oxide for surface sterilization, to a biologically clean (bioclean) environment for reassembly. The operations in this environment are indicated at the right of figure 12.6.

A bioclean facility is an enclosed room with controlled access through an airlock. All air entering the room is first passed through a filter which is designed to remove more than 99.97 percent of all particles larger than 0.3 micron in diameter. The air is changed at least 20 times an hour, using what is called a laminar-downflow type of circulation. The clean air enters through a number of gratings arranged uniformly over the ceiling and it is sucked out at a constant rate, through corresponding gratings in the floor. The air pressure in the room is always maintained slightly higher than outside, so there is no danger of unfiltered air coming into the room. All surfaces are kept free from dust by frequent vacuum cleaning or by wiping.

Before entering the bioclean facility, persons must have their shoes cleaned thoroughly by mechanical brushes. They then scrub their hands and wrists with a germicidal detergent, and don sterilized protective clothing consisting of a hood, surgeon's mask, coverall, gloves, and bootees. Entry into the work area is by way of an airlock, as mentioned above.

By taking adequate precautions, the microbial contamination accumulating on surfaces in a bioclean room is kept to about 100 or less viable organisms per square foot. This number may be compared with an average of several thousand microbes per square foot in normal circumstances.

In the bioclean environment, the subsystems are reassembled once more, after being checked. In order to prevent recontamination of the capsule after the terminal dry-heat sterilization, it is sealed into a canister, which

serves as a biological barrier. The canned capsule is then subjected to the final heat-sterilization treatment designed to reduce the probability of contamination to the desired level, say one part in a thousand. After further tests, the system is ready for launch. The sterile capsule is released from the canister when the vehicle is close to its destination. It must be sufficiently far away, however, to eliminate any possibility that the canister, which has become contaminated on its exterior by handling, will impact on the planet.

The duration of the terminal cycle of heat sterilization will depend on the bacterial load of the capsule and canister. This quantity cannot be determined accurately, but methods are being developed, based on observations of bacterial counts made on exposed surfaces. At one time it was thought that there would be about 10^8 micro-organisms on a space vehicle prior to the final sterilization. It now appears that it may be possible, without too much effort, to reduce the load to some 10^4 or 10^5 organisms. If this is so, the length of the terminal sterilization cycle may be reduced considerably.

In the United States much effort has been devoted to the development of spacecraft components which are not adversely affected by heating at temperatures up to 145° C for several days and by exposure for long periods to a mixture of moist ethylene oxide and Freon-12. Progress has been very satisfactory and there is confidence that a fully heat-sterilizable instrumented capsule can be produced in time for the first opportunity to land a space vehicle on Mars.

It may be noted that the sterilization requirement has been extended to include even rocket fuels, in case retrorockets are required to slow down the space vehicle for landing on Mars. If the fuel is consumed completely, contamination would be no problem because any micro-organisms present would be destroyed by the heat developed. On the other hand, there is a possibility that some fuel will remain or the capsule may crash on the planet's surface instead of making a soft landing. Hence, the rocket fuel must be treated as one of the components of the vehicle capable of withstanding high temperatures.

To judge from published reports, the U.S.S.R. does not have a program for the production of space-vehicle components that can withstand the temperatures desirable for dry-heat sterilization. According to A. K. Astfayeva and her coworkers, whose work on sterilization by radiation was mentioned earlier:

> the complexity of the problem [of spacecraft sterilization] lies in the great number of highly complicated and automatically operating devices manufactured from an immense number of materials which differ in their properties. These features dictate that sterilization of the entire spaceship cannot be performed by using some uniform method of sterilization.

Apparently, the Russians plan to sterilize the space-vehicle components by a procedure that is most suitable for each component; namely, ethylene oxide gas for the exterior and either heat or ionizing radiation for the interior. After complete assembly, the surface of the system will be sterilized with ethylene oxide. "The application of the mentioned methods of sterility and the observation of several measures [which are not described] providing for the establishment of aseptic conditions during assembly of apparatuses and devices, and the final sterilization with the OB [ethylene oxide-methyl bromide] gas mixture may be regarded as an adequate guarantee for the sterility of the spaceship." Some doubt is felt in the United States concerning the adequacy of ethylene oxide for the sterilization of mating surfaces between components of a space vehicle. Unless dieoff, especially of spores, during the flight to Mars results in the complete destruction of organisms on such surfaces, it is not certain that the probability

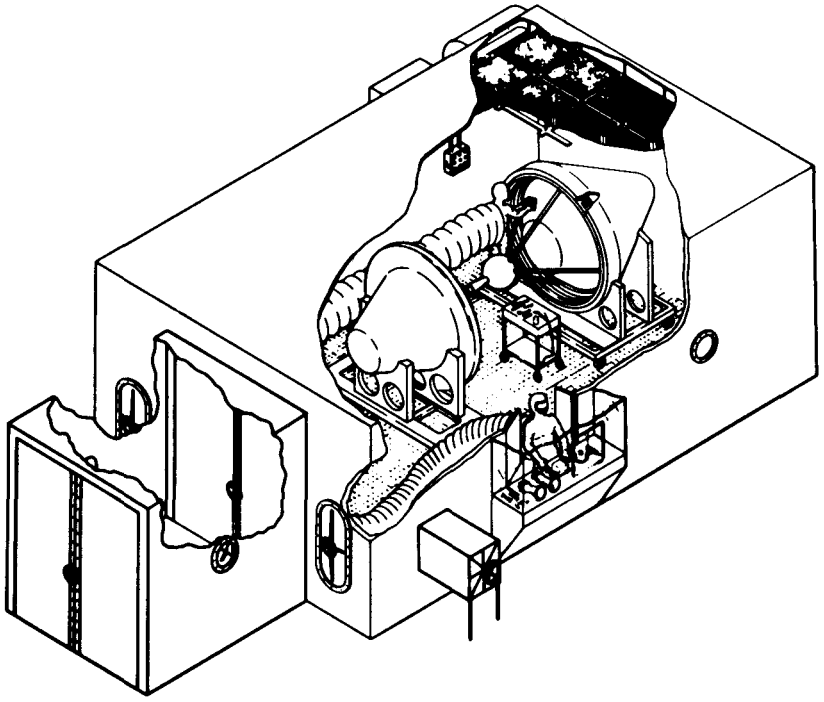


FIGURE 12.7. Representation of a possible method of sterile assembly of spacecraft. (Courtesy of O. W. Nicks and J. R. Miles.)

of a viable organism being aboard the spacecraft will be as low as one in a thousand.

Sterile Assembly

Although it is not being considered for use in the near future, long-range plans envisage the possibility of completely sterile assembly of subsystems. Such a procedure would perhaps permit less severe terminal sterilization than is now contemplated. Furthermore, if there were components which could not withstand high temperatures but could be sterilized by radiation, they could be inserted after heat sterilization of the remainder of the spacecraft. Another advantage would be that replacement of a component which failed at a late stage would be very much simpler than it is under present conditions.

A possible approach to sterile assembly is shown in figure 12.7. This is an extension of the glove-box concept in which workers perform operations in a small sterile chamber with their hands in gloves sealed into one wall of the chamber. In figure 12.7, the operators are completely enclosed in tunnel suits, as they have been called. Within these suits, workers have access to outside air, but they have sufficient mobility to permit them to assemble sterilized components in sterile surroundings.

Avoidance of Impact With Mars

In addition to requiring a low probability of a viable organism remaining aboard a

sterilized landing capsule, the COSPAR specifications, it will be recalled, stipulate a very small probability that an unsterilized space vehicle will suffer accidental impact with Mars. The problem here is largely that of providing accurate guidance for the spacecraft on its way to the planet. In general, one or two midcourse corrections to the initial trajectory may be required, after the characteristics of the latter have been determined. Trajectory calculations can be made with great accuracy, but the control of the rocket motors which perform the necessary corrections is less precise. With developments in technology, however, the situation is continually improving.

The ability to make adjustment to the trajectory depends, of course, on the transmission of radio signals to and from the spacecraft. If the power source should fail, as it apparently did on the Russian Mars probe Zond 2, then nothing can be done to prevent an accidental impact with the planet. It may well turn out that the critical factor in avoiding contamination of Mars is not sterilization of landing capsules but prevention of impact by flyby and orbiting vehicles.

FOR FURTHER INFORMATION

AFASTAYEVA, A. K.; ET AL.: Methods for Spacecraft Sterilization. Vol. V of Life Sci. Space Res., North-Holland Publ. Co., 1967, p. 38.

CRAVEN, C. W.; STERN, J. A.; AND ERVIN, G. F.: Planetary Quarantine and Space Vehicle Sterilization. Jet Propulsion Laboratory Tech. Rept. no. 32–1276, 1968. (Reprinted from Astronaut. Aeron., vol. 6, no. 8, 1968, pp. 18–48.)

EXOTECH, INC.: Study of Analytical Techniques in Planetary Quarantine. Final Report Contract NASw–1550 for NASA Office of Biosciences, 1967.

HALL, L. B.; AND BRUCH, C. W.: Procedures Necessary for the Prevention of Planetary Contamination. Vol. III of Life Sci. Space Res., North-Holland Publ. Co., 1965, p. 48.

HALL, L. B.: The Importance of Sterilization Techniques in Space Exploration. Vol. VI of Life Sci. Space Res., North-Holland Publ. Co., 1968.

HAWRYLEWICZ, E. J.; ET AL.: Probability of Growth of Viable Organisms in Martian Environments. Vol. VI of Life Sci. Space Res., North-Holland Publ. Co., 1968.

HOROWITZ, N. H.; SHARP, R. P.; AND DAVIS, R. W.: Planetary Contamination. I: The Problem and the Agreements. Science, vol. 155, 1967, p. 1501.

LIGHT, J. O.; ET AL.: A Discussion of the Planetary Quarantine Constraints. Vol. V of Life Sci. Space Res., North-Holland Publ. Co., 1967, p. 7.

LORSCH, H. G.: Biocontamination Control. Space/Aeronautics, Nov. 1966, p. 89.

LORSCH, H. G.; AND KOESTERER, M. G.: Biological and Engineering Aspects of Spacecraft Sterilization. Vol. V of Life Sci. Space Res., North-Holland Publ. Co., 1967, p. 22.

MURRAY, B. C.; DAVIES, M. E.; AND ECKMAN, P. K.: Planetary Contamination. II: Soviet and U.S. Practices and Policies. Science, vol. 155, 1967, p. 1505.

NICKS, O. W.; AND MILES, J. R.: Some Developments in Sterilization for Planetary Probes. Vol. IV of Life Sci. Space Res., Spartan Books, 1966, p. 176.

PACKER, E.; SCHER, S.; AND SAGAN, C.: Biological Contamination of Mars. II: Cold and Aridity as Constraints on the Survival of Terrestrial Micro-organisms in a Simulated Martian Environment. Icarus, vol. 2, 1963, p. 293.

PITTENDRIGH, C. S.; VISHNIAC, W.; AND PEARMAN, J. P. T., EDS.: Biology and the Exploration of Mars. Nat. Acad. Sci.-Natl. Res. Council Publ. 1296, 1966.

SAGAN, C.; LEVINTHAL, E. C.; AND LEDERBERG, J.: Contamination of Mars. Science, vol. 159, 1968, p. 1191.

SPACECRAFT STERILIZATION TECHNOLOGY. NASA SP–108, 1966.

XIII

Exploration of Mars With Spacecraft

GENERAL PRINCIPLES

Types of Spacecraft

APART FROM THE PROBLEM of whether or not life exists, or has existed, on Mars, there are many questions relating to the physical characteristics of the planet which cannot be answered by observations made from Earth. It is only by the use of spacecraft, which can approach or land on the surface of Mars, that reliable information can be obtained on such matters as the composition and structure of the atmosphere, the nature of the polar caps, the composition of the surface, and the cause of the various changes in appearance that occur from time to time.

The main types of spacecraft to be used for the study of Mars are landers, orbiters, and flyby vehicles, as indicated in the preceding chapter. In principle, each type of vehicle can be manned or unmanned. There are many difficult problems to be solved, however, before men can be sent to Mars and returned safely to Earth. The discussion in this chapter, therefore, will be concerned with unmanned spacecraft carrying various instruments for the exploration of Mars at close range.

In addition to the three kinds of spacecraft referred to above, an intermediate type should be mentioned. To determine the properties of the Martian atmosphere, which must be known before a vehicle capable of making a soft landing on the surface can be designed, it may prove desirable to use one or more atmospheric probes. Such a probe, or hard lander, would be a small spacecraft, 2 or 3 feet in diameter, which would be launched from a flyby or orbiter spacecraft, and would crash on the surface of the planet. The probe would carry instruments for measuring the density, pressure, and possibly the temperature at various altitudes as it descended through the Martian atmosphere.

Interplanetary Trajectories

No matter what type of spacecraft is used in the exploration of Mars, it must be launched from Earth into a suitable interplanetary trajectory by means of a rocket vehicle. So that the spacecraft may carry the maximum amount of instrumentation, the

trajectory should be achieved with the minimum expenditure of energy. The extra energy available from the rocket engines can then be utilized for the payload of instruments.

The ideal trajectory requiring a minimum of energy for a journey between two planets is called the Hohmann transfer ellipse, named after the German engineer W. Hohmann who described it in 1925. The simplest, although unreal situation, is that in which two planets, Earth and Mars, for example, travel in circular orbits lying in the same plane. The trajectory requiring the least expenditure of energy is the one in which the spacecraft leaves one planet in a direction tangential to its orbit and arrives at the other planet tangentially to the orbit of the latter. The resulting trajectory is the Hohmann transfer ellipse (which is actually half an ellipse), and it is shown in figure 13.1.

The transit time between Earth and Mars for an idealized Hohmann trajectory is readily calculated to be 260 (Earth) days. The sidereal period of Mars, which is the time

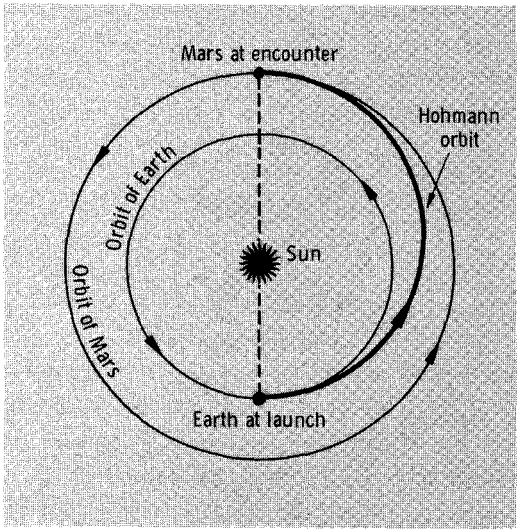


FIGURE 13.1. An ideal Hohmann transfer ellipse.

taken to complete its orbit (360 degrees) about the Sun (p. 37), is 687 days. Consequently, Mars moves through an angle of (260/687) $\times 360 = 136$ degrees about the Sun during the time the spacecraft is in transit between Earth and Mars. It can be seen from figure 13.2, therefore, that for a Hohmann transfer ellipse to be possible, the relative positions of Earth and Mars at the time of launch must be such that the angle Earth-Sun-Mars is $180 - 136 = 44$ degrees.

A situation of this kind, referred to as a "Mars opportunity," arises only once in every Martian synodic period (the time between successive Earth-Mars oppositions) of 780 days, or 2 years and 50 days, on the average. From the data in chapter III it can be shown that the ideal opportunity when Earth is 44 degrees behind Mars in solar orbit, occurs about 50 days before each opposition. Since the journey takes 260 days, encounter between the spacecraft in an ideal Hohmann trajectory and Mars will occur some 210 days after opposition.

The location of Earth at the time of encounter is of interest because of the transmission of information by radio from the spacecraft in the vicinity of Mars. Earth travels a complete orbit around the Sun in 365 days, so, during the period of 260 days in which the spacecraft is in transit, Earth will have traversed an angle of (260/365) $\times 360 = 256$ degrees about the Sun. This is $180 + 76$ degrees, and so the Earth-Sun-Mars angle when the spacecraft is near Mars would be 76 degrees, as shown in figure 13.2. The distance from the spacecraft (or Mars) to Earth at the time of encounter after an ideal Hohmann trajectory from Earth to Mars would then be about 200 million kilometers (125 million miles).

The orbits of Earth and, especially, of Mars deviate from a circle and, furthermore, the two orbits are not in the same plane.

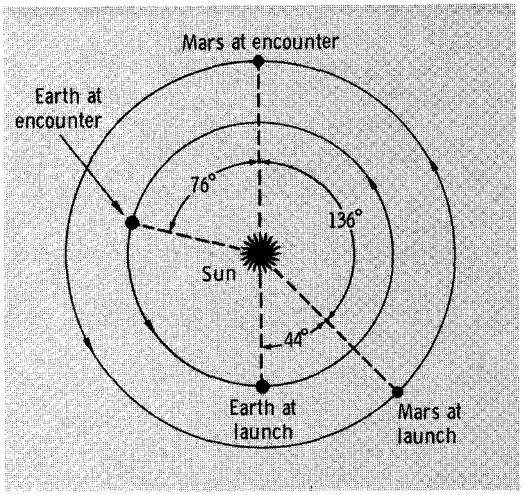


FIGURE 13.2. Relative positions of Earth and Mars at launch and encounter for Hohmann transfer ellipse.

Consequently, the calculations based on the Hohmann transfer ellipse represent an ideal situation which cannot be realized in practice. Nevertheless, the general conclusions give a rough indication of the launch and encounter conditions. In particular, trajectories requiring the lowest energies occur within a launch period of about 2 or 3 months before each Martian opposition. In 1969, for example, the opposition date is May 31, and the launch period is around February and March. In 1971, with the opposition on August 10, the ideal launch period would be roughly May to June.

As a consequence of the fairly large ellipticity of the Martian orbit, the distance from Earth to Mars varies from one opposition to another. As seen in chapter III, favorable oppositions, when Mars and Earth are at their closest, occur every 15 or 17 years. At these times it requires less energy than at other oppositions to transfer a given mass from Earth to Mars. An approximate indication of the relative minimum injection energies at various oppositions from 1967 through 1980 is given in figure 13.3. It is seen that in 1969, and especially in 1971, when a favorable opposition occurs (fig. 3.7), the energy requirements will be less than in other years.

In the ideal Hohmann trajectory, the spacecraft travels exactly halfway around the Sun. The angle around the Sun swept out by the spacecraft is then exactly 180 degrees. In practice, the actual trajectories from Earth to Mars fall into two categories: type I in which the angle around the Sun is less than 180 degrees, and type II in which the angle is greater than 180 degrees (fig. 13.4). In general, type I trajectories have shorter flight times than those of type II.

The type of trajectory selected for a Mars mission depends on many circumstances. Among these are the flight directions when the spacecraft is launched from Earth and when it reaches Mars. For safety reasons, spacecraft launched from the Eastern Test Range, Cape Kennedy, Fla., as would be the case for Martian missions from the United

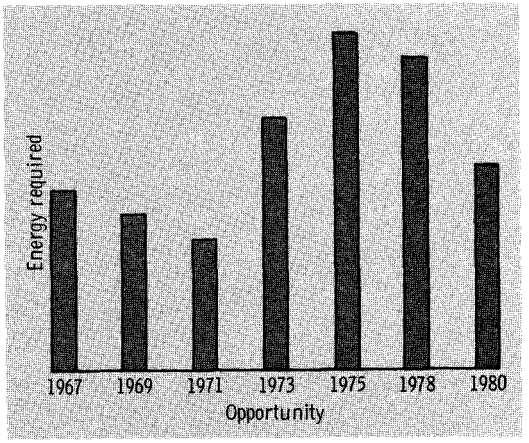


FIGURE 13.3. Minimum injection energies at various Martian oppositions. (After C. Sagan and J. W. Haughey.)

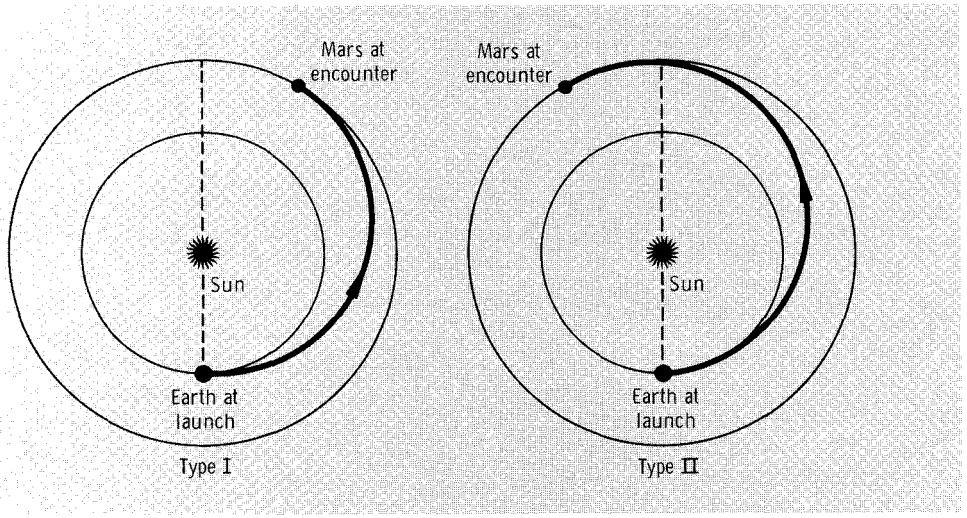


FIGURE 13.4. Type I (left) and Type II (right) trajectory from Earth to Mars.

States, must make their initial flight over uninhabited ocean areas. The launchings are thus restricted to a southeasterly direction. Furthermore, the angle of approach to Mars must be considered because this will determine the area of the planet that can be observed by a flyby vehicle or the final location of a lander. Another factor is the requirement that radio communication be maintained between the spacecraft when it is near Mars and the receiving stations of the U.S. Deep Space Network on Earth.

The actual trajectory of a spacecraft from Earth to Mars thus represents a compromise among many factors. It is very rarely, therefore, that the trajectory approaches the one of minimum energy for the given opportunity. Nevertheless, the launch period is always close to that indicated by the ideal calculations given above. The flight path may be shorter (type I trajectory) or longer (type II trajectory) than the ideal value, but the energy requirement is always larger.

As the result of many calculations, it has been decided that type I trajectories are preferable for the 1969 opportunity. Because the minimum energy requirement for the ideal trajectory, as indicated in figure 13.1, is relatively low, it is possible for the actual trajectory to depart considerably from the ideal. The flight times from Earth to Mars are much shorter than the Hohmann value, being from about 130 to 160 days. A consequence of the short travel time is that the distance from the spacecraft to Earth at the time of encounter will be only some 100 million kilometers (62.5 million miles). This will greatly facilitate radio communications between the spacecraft and Earth. A factor contributing to the relatively short distance between Earth and Mars at the time of a spacecraft encounter with Mars in August 1969 is that the planets are then in the region where their orbits come closest together (fig. 3.7).

Trajectory Injection

For a spacecraft to follow a trajectory that will lead to an encounter with Mars, it must escape from control by Earth's gravitational

field and become a satellite of the Sun. The first requirement is, therefore, that the geocentric injection velocity, that is, the velocity with respect to Earth, must exceed the terrestrial escape velocity. This is about 39 500 kilometers (24 500 miles) per hour at the altitude attained by the rocket vehicle.

After attaining the escape velocity, the spacecraft must be injected into a trajectory that is an orbit about the Sun. Because the orbit of Mars lies beyond that of Earth in the solar system (fig. 3.1), the heliocentric velocity of the spacecraft, that is, its velocity with respect to the Sun, must exceed the orbital velocity of Earth. To achieve this, the spacecraft is launched in the same direction as Earth is traveling in its orbit. The injection velocity of the vehicle is then added to Earth's orbital velocity. Both the injection velocity and the point in space at which injection into solar orbit takes place depend on the desired trajectory to Mars.

Two general procedures are available for injecting a spacecraft into an interplanetary (Earth to Mars) trajectory. In the first, the vehicle is launched into an Earth orbit, called a parking orbit or, better, a coasting orbit, at an altitude of about 160 kilometers (100 miles). The rocket engine is then shut off and the space vehicle coasts for a short time in Earth orbit. Just before the required injection point is reached, the engine is restarted and the speed of the spacecraft is increased until it is sufficient to achieve the necessary velocity for injection at the proper location in space. The second possibility is to use the procedure known as direct ascent, in which the speed of the vehicle is increased continuously until the injection velocity at the required point is attained. When the space vehicle is in its interplanetary trajectory, the rocket engine used to launch it is jettisoned. Except for a minor adjustment in velocity, the spacecraft continues to Mars without any further application of thrust.

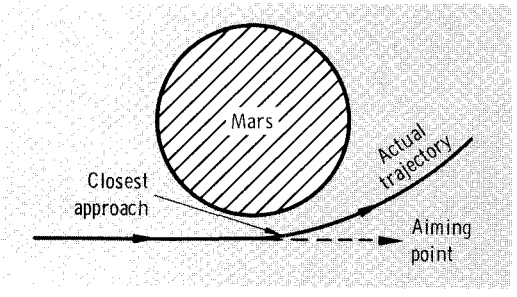


FIGURE 13.5. Aiming point and Mars trajectory.

The two launch modes described above are often referred to as "two-burn" and "one-burn," respectively. In the former, the rocket engine, actually the upper stage, as will be seen in due course, is ignited, or burns, twice, first to attain the proper coasting orbit and again, after a lapse of time, to achieve the required injection velocity. In the one-burn, direct-ascent mode, however, the engine is ignited only once. The two-burn mode of injection into an interplanetary orbit is somewhat more flexible and it makes better use of the rocket engine thrust. On the other hand, the one-burn injection is more reliable because it does not involve restarting a rocket engine in space.

The trajectory of a spacecraft is designed to bring the vehicle reasonably close to Mars, but not close enough to cause it to impact on the planet. The "aiming point" near Mars is farther from the planet's surface than is the distance of closest approach, because the spacecraft is attracted by Martian gravity when it is near the planet, and the trajectory is changed to some extent (fig. 13.5). Allowance for the gravitational attraction of Mars is made, of course, in calculating the trajectory of the spacecraft. Because of inevitable minor deviations in injection close to Earth, the actual trajectory of about 300 to

500 million kilometers (180 to 300 million miles) from Earth to Mars will deviate somewhat from the one desired. To compensate for such deviations, a so-called "midcourse" trajectory correction maneuver is performed, generally a few days after launch and long before the spacecraft reaches the middle of its long journey to Mars.

The interplanetary trajectory of a spacecraft is established as precisely as possible from tracking data obtained during the first week or so of the flight. The change in velocity, in both speed and direction, required to bring the trajectory close to the desired one is then calculated. Commands are transmitted from Earth to the spacecraft for it to take up a particular orientation (attitude) in space and to fire a rocket motor on the spacecraft for a certain specified time. These operations constitute the midcourse maneuver. If necessary, a second maneuver can be performed at a later stage in the trajectory. Details of the change made in December 1964 in the trajectory in the Mariner IV mission to Mars are described later.

SCIENTIFIC OBSERVATIONS FROM SPACECRAFT

Observations From Atmospheric Probes

A method has been proposed for determining the structure of the Martian atmosphere from observations made with a small probe ejected from a flyby or orbiter spacecraft. All that is required, in principle, is an accelerometer, which is a simple instrument for measuring the rate of change of speed of the probe as it descends through the atmosphere to the surface of the planet. Because of the aerodynamic resistance (drag) of the atmosphere, the speed of the probe decreases. The rate of decrease in speed, generally referred to as an acceleration although it is actually a deceleration, is related to the density of the atmosphere. In the simple case of a spherical probe, the density ρ at a particular altitude may be represented by

$$\rho = \text{constant} \times \frac{a}{V^2}$$

where a is the acceleration of the probe and V is its speed at the given altitude. In this expression, the constant, which is a specific property of the probe, can be calculated from terrestrial measurements.

In order to utilize the foregoing equation to determine the structure of the atmosphere, it is necessary to obtain the values of a and V for a known altitude h over a range of altitudes. Of the three quantities—a, V, and h—only a is measured directly by the accelerometer system, generally consisting of three accelerometers oriented at right angles to one another. From the manner in which the acceleration varies with time, however, it is possible to calculate the speed V at a height h at any instant. To do this, it is necessary to know the speed, altitude, and path angle, that is, the angle of the direction of motion of the probe relative to the Martian horizon, at the time the probe enters the atmosphere. These quantities can be derived from tracking observations made on the space vehicle at the time the probe is ejected.

The calculations based on the acceleration data yield the density of the atmosphere as a function of height above the Martian surface. It is then a simple matter, by using the familiar barometric equation, to find how the pressure of the atmosphere changes with altitude. Thus, observations from an atmospheric probe can provide valuable information on the variation with altitude of the density and pressure in the Martian atmosphere. Incidentally, although the equation given above refers to a spherical probe, the general principles described are applicable to

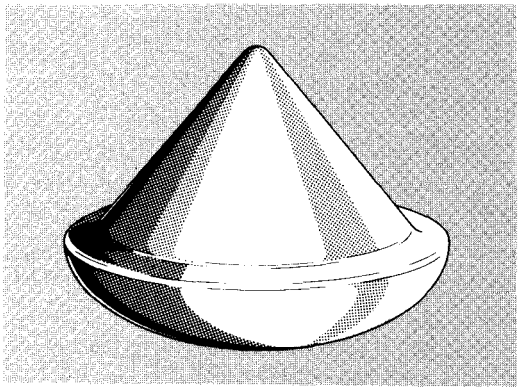


FIGURE 13.6. A Martian atmospheric probe.

probes of other shapes. There are some advantages, for example, in using an atmospheric probe like that depicted in figure 13.6.

As the probe approaches the surface of Mars, its speed is considerably lower than at the instant of its separation from the carrier vehicle and entry into the atmosphere. When the actual speed is less than about one-fifth that at entry, the calculations, especially of the speed, may be very much in error. The procedure, which is quite satisfactory at high altitudes, becomes unreliable at lower altitudes when the actual speed is less than the speed of sound in the surrounding atmosphere. On the other hand, direct measurements of the normal atmospheric temperature and pressure, which are not feasible at high speeds, become possible when the probe has slowed down. The temperature can be determined with a small platinum resistance thermometer and the pressure with a device which depends on the effect of the atmosphere on a thin metal diaphragm.

From the variation of the temperature and pressure with time, the temperature and pressure at any altitude can be evaluated. Furthermore, just as the pressure can be calculated from the density, as mentioned above, the density can be derived from the pressure and altitude. Thus, pressure, temperature, and density can be obtained as functions of altitude. From the former three quantities, the mean molecular weight M of the atmospheric gases at each altitude can then be determined from the equation

$$M = \text{constant} \times \frac{\rho T}{p}$$

where p is the pressure and T is the absolute temperature corresponding to the density ρ at any given altitude. The constant in this expression is a well-known quantity and is independent of the nature of the atmospheric gases. In writing this expression for the mean molecular weight, it is assumed that the gases behave in an ideal (perfect) manner. At the low pressures existing in much of the Martian atmosphere, this assumption will cause little error.

A suggestion has been made for the inclusion of a compact form of mass spectrometer (p. 231) in an atmospheric probe for use near the surface when the speed has been reduced. In this event, the individual molecular weights of the constituent gases of the atmosphere could be obtained directly. A mass spectrometer would not only identify the gases present, it would also indicate their relative amounts at different altitudes.

Information concerning some of the gases in the atmosphere of Mars can be obtained in another way. When the probe is traveling at high speed, the ambient atmosphere is heated by the shock wave generated by the rapid motion. As indicated above, temperature measurements then bear no relation to the normal atmospheric temperature. But the heated gases in the shock layer are stimulated to emit characteristic radiations which might be utilized to identify the gases present.

If the atmosphere of Mars consists mainly of carbon dioxide and some nitrogen, as ap-

pears reasonable, then the radiation should contain the violet emission band of the cyanide (CN) radical, with a maximum at a wavelength of about 4200 Å. From the variation of the intensity of this band with the speed of the probe, it should be possible to obtain a fairly good indication of the proportions of carbon dioxide and nitrogen in the atmosphere. Additional data relating to the interpretation of the CN bands could be derived from observations on the Swan bands of the diatomic carbon molecule (C_2) and of the nitrogen molecule ion (N_2^+). For such measurements, a simple optical radiometer, capable of determining radiation intensities at certain specified wavelengths, has been developed for carriage aboard an atmospheric probe.

There is another aspect of the heating caused by the shock wave ahead of the rapidly moving probe that must be taken into consideration. The high temperature causes the gases to become ionized and this results in the formation of what is called a plasma sheath, consisting of ions and electrons. Such a sheath blocks the passage of radio signals and so leads to what is known as a radio blackout. Hence, information obtained from the accelerometer and radiometer systems during the blackout period must be stored on tape in the probe. It is then read out (or played back) when the speed has decreased and the plasma sheath has dispersed. Signals from the probe would be received by the vehicle from which it was launched and then retransmitted to Earth.

Observations From Flyby Vehicles

The most direct observations that can be made from the flyby spacecraft are by means of television cameras. Remarkable pictures of the Martian surface were taken by the Mariner IV spacecraft in 1965. Although the cameras were between 16 900 and 11 900 kilometers (10 500 to 7400 miles) from the surface and the data were transmitted over distances of more than 220 million kilometers (136 million miles), many of the pictures showed excellent detail (figs. 6.17 and 6.18). With improvements in technique and a close approach to the planet, even better photographs of the surface of Mars should be obtained in 1969.

Determination of the local temperatures at different points on the Martian surface can be made by means of an infrared radiometer on a flyby vehicle. If the surface is obscured by clouds which are opaque to the infrared radiation from the surface, the radiometer will give the temperature at the tops of the clouds. The principle is the same as that described in chapter VI for estimating the surface temperature of Mars from observations of infrared emission made from Earth.

The total energy in one or, better, two wavelength regions in the infrared emitted by the surface of Mars is measured by the radiometer, and from the results the temperature of the emitter can be calculated. The wavelength regions selected are those in which the gases in the Martian atmosphere, particularly carbon dioxide, absorb infrared radiation to a negligible extent (see below). Because infrared radiation measurements from a flyby spacecraft can be made at relatively close range of a few thousand kilometers, they will be very much more reliable than those obtained from terrestrial observations. Furthermore, the radiometer on a space vehicle can be directed at specific areas, thus permitting local temperatures to be determined with much greater precision than is possible from Earth.

In addition to measurements by means of a radiometer of the infrared radiation emission from the surface, observations can be made with a spectrometer of the infrared

radiation absorbed by the Martian atmosphere at specific wavelengths. As seen in chapter V, the infrared spectrum of Mars, as observed on Earth, provided evidence for the existence of carbon dioxide and water vapor in the atmosphere of the planet. The infrared absorption spectrum obtained from a spectrometer on a flyby spacecraft will, of course, give much better data about these gases and the possible variations in the quantities at different localities.

The infrared absorption spectrum may also indicate the presence of other gases, such as ammonia, methane, oxides of nitrogen, cyanogen, and hydrogen cyanide, which may be present in small quantities in the atmosphere of Mars. The amounts of such gases would be too small to be detectable from Earth, but their existence would be important in determining whether the Martian atmosphere is primarily oxidizing or reducing in nature. This information, as well as the departure, if any, of the composition from thermodynamic equilibrium would have a bearing on the possible presence of life on the planet.

Of particular interest would be a check on the C–H band reported by W. M. Sinton at a wavelength of 3.43 μ (p. 220). Fairly large (polyatomic) molecules in the atmosphere could be detected by their infrared absorption. If present, they would suggest the occurrence of biochemical processes. Such molecules also absorb ultraviolet light and so they might have some influence on the intensity of this radiation that reaches the surface of Mars.

Nitrogen molecules, which may exist in the Martian atmosphere, cannot be determined with an infrared spectrometer because they do not absorb infrared radiation to a significant extent. There is a good possibility, however, that nitrogen, as well as other atoms and molecules of interest, could be identified by their ultraviolet spectra. It is known that radiation from the Sun is able to stimulate the emission of their characteristic spectra by some of the constituents of the upper atmosphere of Earth. There is little doubt that a similar phenomenon can occur on Mars. The species that might consequently be detected in the Martian atmosphere and the wavelengths of characteristic lines in the ultraviolet or visible emission spectra are listed in the accompanying table.

Species Detectable by Ultraviolet (or Visible) Emission Spectra

	Formula	*Wavelength (Å)*
Atoms:		
Hydrogen	H	1216
Oxygen	O	1304
Nitrogen	N	1200
Carbon	C	1657
Ions:		
Nitrogen molecules	N_2^+	3914 (visible)
Carbon monoxide	CO^+	4264 (visible)
Carbon dioxide	CO_2^+	2880
Molecules:		
Nitrogen	N_2	1354, 1371
Nitric oxide	NO	2150
Carbon monoxide	CO	2160
Cyanogen	C_2N_2	3876 (visible)

An ultraviolet spectrometer, capable of observing the emission spectrum in the upper layers of the atmosphere of Mars, could be carried on a flyby spacecraft. An important difference in use, apart from the wavelength region being observed, between an ultraviolet (emission) spectrometer and an infrared (absorption) spectrometer is that the latter would be pointed downward toward the surface, whereas the former would look into the atmosphere in a direction parallel to the surface. It would consequently be possible to study the composition of the atmosphere at a series of different altitudes by means of an ultraviolet spectrometer. The manner in which the amounts of the various species present change with altitude can thus be determined.

Another type of instrument that could be aboard a flyby spacecraft is a magnetometer for determining the strength of the magnetic field of Mars. The intrument on Mariner IV indicated that the field at a distance of about 10 000 kilometers (6000 miles) from the surface was indistinguishable from that in interplanetary space. The conclusion drawn, therefore, was that the magnetic field of Mars was either very small or, possibly, nonexistent. It might be of interest, however, to make a more detailed study. This could be done by means of a magnetometer on a flyby vehicle approaching closer to Mars than did the Mariner IV spacecraft.

The use of radio signals from Mariner IV to study the atmosphere and ionosphere of Mars was described in chapter V. This occultation experiment does not require any special equipment. The radio transmitter on the spacecraft, which provides the signals for tracking purposes and for telemetry of information from the instruments, automatically supplies all that is required for the occultation experiment.

Observations From Orbiters

An orbiter can carry the same instruments as a flyby vehicle, but it would have the great advantage of being able to make repeated observations for some time over a large part of the Martian surface. The orbit of a satellite, such as an orbiting spacecraft, remains virtually stationary in space. But because the planet Mars rotates, once in a little over 24 hours, one region after another would come within view of the orbiter. If the orbital plane of the spacecraft is approximately at right angles to the equator, that is, if the vehicle is in a roughly polar orbit, all areas of Mars could be observed. Both daily and seasonal changes in atmospheric composition and surface temperature for different areas of the planet could be studied in this manner.

An infrared radiometer can give the surface or cloudtop temperature, but similar emission measurements in the microwave region can indicate the temperatures below the surface (p. 135). A knowledge of such temperatures at different depths at various locations on Mars and their daily and seasonal variations, if any, would be of great value in understanding the transport of heat below the surface of the planet.

If the spacecraft could carry a radio-transmitter covering a range of wavelengths, as well as an appropriate receiver, it would be possible to make a more precise and detailed study of the Martian ionosphere than by the occultation method. The principle involved is exactly the same as that used in an ionospheric sounder (ionosonde) to investigate Earth's ionosphere. A series of radio pulses of higher and higher frequency are transmitted vertically upward (from Earth's surface) or downward (from a satellite) into the ionosphere, and the times of the return, or echo, signals are observed. From this information the structure of the ionosphere can be derived. An ionosonde transmitter and receiver on a Martian orbiter could be utilized in the same way to investigate the ionosphere of the planet. A possible simplification would be to use a powerful transmitter on Earth to send the incident signals into the ionosphere of Mars, and then to detect the echo signals with a receiver on the orbiter.

An orbiting space vehicle could be utilized to study the topography of the Martian surface in some detail. For example, a radar altimeter would give the vertical distance of the orbiter from the surface at various locations, and from these data differences in altitude, such as between adjacent bright and dark areas, could be computed. Another approach is to use an infrared spectrometer,

pointing downward, to determine the amount of carbon dioxide in the atmosphere. This should be greater over the low-lying areas than it is over the highlands. Finally, advantage could be taken of the fact that the radio signal from an orbiter to Earth is occulted by Mars once in every orbit of the spacecraft. The elevations of the areas grazed by the radio beam can then be calculated from the observed times of entry into and emergence from occultation. These areas will vary from one orbit to the next.

It was seen in chapter VI that the polarization characteristics of the sunlight reflected (or scattered) from Mars have been used to provide information about the surface structure and, possibly, about the nature of the surface material. A polarimeter on an orbiter would yield much better data than can be obtained from Earth. The instrument would be much closer to the Martian surface, and it would be possible to measure the polarization over a much greater range of phase angles. Such data might indicate the composition and roughness of the surface materials at different locations on the planet. Areas which showed distinct seasonal changes in polarization might perhaps be those selected as sites for later landing missions.

Observations From Landers

The main use of instruments deposited on the surface of Mars by a lander would be to seek for evidence of life, past or present, or of substances of biological interest. The kinds of experiments which might be performed in this connection were described in chapter XI. In addition, observations could be made on the atmosphere and soil, as indicated in the description of the Automated Biological Laboratory in that chapter. The brief review here will be concerned mainly with instruments which could be developed in compact form and of small mass, and which could be used to study the atmosphere and soil of Mars.

A lander would descend to the surface at a relatively slow speed to avoid damage to its payload of instruments. Consequently, it would be possible to make direct measurements of the temperature and pressure of the atmosphere during the descent. The procedures would be similar to those already described in connection with the later stages of the fall of an atmospheric probe. Furthermore, the composition of the atmosphere could be determined by a gas-solid chromatograph, based on absorption of the gases by a solid, or a mass spectrometer, or a combination of both. All these instruments would continue to operate and record atmospheric temperature, pressure, and composition, and their diurnal and seasonal variations, after the lander had settled on the surface of Mars.

There are difficulties in determining the amount of water vapor in the atmosphere by chromatography. A simple device has been developed, however, for making such measurements. It consists essentially of a capacitor with two plates of alumina (Al_2O_3). The impedance of this capacitor, which is readily measured, depends on the amount of water vapor present between the plates.

Several methods have been proposed for determining the elements present in the Martian soil. These include observations of atomic spectra, of X-ray fluorescence excited by fast electrons, of the scattering of alpha particles, and of gamma rays produced by bombardment with neutrons. The latter two are of special interest, because compact instruments have been developed. Another possibility is to make use of an X-ray diffractometer, as comparison with diffraction patterns of terrestrial minerals may indicate what inorganic substances are present on the surface of Mars.

319-889 O—69——9

The scattering of alpha particles, that is, helium nuclei, emitted by a radioactive source has been employed to identify elements on the lunar surface and to determine their relative amounts. A narrow, collimated (parallel) beam of alpha particles is allowed to fall on the material under examination, such as the surface of Mars. The alpha particles scattered, or reflected, through a large angle (about 160 degrees) fall on a detector which gives the number of particles as a function of their energy. For each element heavier than aluminum (atomic weight 27), the cutoff energy, beyond which no scattering occurs, is typical of the element. Hence, the scattering energy spectrum can be used to determine the elements present in the surface material.

In addition to the alpha particles that are scattered, some are captured by the nuclei of certain elements, and protons (hydrogen nuclei) of characteristic energy are emitted. These would be observed by separate proton detectors and would provide information concerning the lighter elements in the target material exposed to the alpha particles. It has been claimed that the alpha-scattering instrument, with alpha-particle and proton detectors, is capable of identifying all elements, except hydrogen, helium, and lithium, present in amounts larger than 1 atom percent in the sample under examination.

The method of neutron bombardment depends on the fact that when neutrons of high energy interact with the nuclei of different elements, characteristic types of gamma rays are produced. From observations on these gamma rays, the elements present can be identified. The gamma rays produced by the action of neutrons fall into three categories: inelastic scattering, radiative capture, and activation gamma rays. One, two, or all three types may be generated with any particular element, but they can be recognized by their appearance at different times. The high-energy (fast) neutrons required for this analytical procedure are obtained by allowing accelerated deuterons (deuterium nuclei) to bombard a tritium target.

When a nucleus of an element interacts with a neutron of high energy, some of this energy may be converted into internal energy of the nucleus. This process is called inelastic scattering of the neutrons. Within an extremely short time, the extra (internal) energy of the nucleus is emitted as a gamma ray. The inelastic scattering gamma ray thus produced is characteristic of the element from which it is formed. By measuring the energy of the gamma ray, the element can be identified. The procedure is applicable to several common elements found in terrestrial rocks, including oxygen, aluminum, silicon, magnesium, sodium, potassium, and iron.

After a fast neutron is slowed down, it may be captured by a nucleus. As a consequence of this process, radiative capture gamma rays are usually emitted from the resulting nucleus. These rays are also characteristic of the element whose nuclei have captured the slow neutrons. Radiative capture gamma rays can be distinguished from inelastic scattering rays by the fact that the former are emitted a significant fraction of a second after the latter.

Activation gamma rays are generated in some cases when the nucleus resulting from the capture of a neutron is radioactive. Gamma rays of specific energies are then emitted, with decreasing intensity, for some time at a rate dependent on the half-life of the radioactive species. Activation gamma rays are thus readily distinguishable from the other two types that might result from interaction with neutrons. It is possible to design an instrument which will detect separately the three kinds of gamma rays, but for practical purposes, such as for use on the surface of Mars, it appears advisable to restrict the meas-

urement to inelastic scattering gamma rays.

The alpha-particle scattering, neutron-interaction, and other analytical methods referred to above, can only identify the elements, or, in the case of X-ray diffraction, minerals that may be present in the soil. But special interest is attached to organic compounds because of their biological significance. For the detection and analysis of such compounds, the procedure involving pyrolysis, followed by gas-liquid chromatographic separation and mass spectrometry, as described in chapter XI, has been suggested. The method could be applied directly to a sample of soil, without prior treatment. Analysis of separate fractions obtained by extraction of the soil with solvents would be more reliable, but the equipment required would be more complicated.

There are several other instruments which could be carried by a lander to the surface of Mars. They could be employed to determine surface hardness, ground motion, wind velocity and direction, light intensity, ultraviolet radiation, cosmic rays, and so on. Finally, just as television cameras have proved useful in studying the surface of the Moon, they could be utilized to provide information on the topography in the region of a vehicle landed on Mars.

THE MARINER IV MISSION

The Mariner IV Spacecraft

The first information from Mars to be secured by means of a spacecraft was obtained from Mariner IV in July 1965. The Mariner IV space vehicle consisted of an octagonal framework to which were attached four large solar-cell panels, as seen in figure 13.7. Most of the octagonal section was occupied by electronic and other equipment for radio communication to and from Earth and for storing data from the television camera and scientific instruments. Bottles of nitrogen gas required for control of the orientation (attitude) of the spacecraft were also contained in this section. The rocket motor used in the midcourse maneuver was located in one face of the octagon. Six of the other faces were fitted with louvers for thermal control. By opening and shutting automatically, as needed, they helped to maintain the equipment in the interior at a fairly constant temperature.

Some of the scientific instruments carried by Mariner IV were mounted on top of the octagonal structure, while others were attached to a mast extending from it. At the top of this mast was the low-gain omnidirectional antenna, whereas at the bottom, supported on the octagonal base, was a high-gain, directional (dish-shaped) antenna. The television camera was at the bottom of the base, as shown in the lower illustration in figure 13.7. It was mounted on a scan platform which could be rotated so as to permit the camera to view the surface of Mars when Mariner IV was in the vicinity of the planet.

The solar-cell panels were covered with a total of more than 28 000 small cells. When exposed to the Sun, these cells took up energy from the solar radiation and converted it into electric power. The electricity generated in this manner was utilized for operating instruments, data-storage equipment, radio receiver and transmitter, etc., on the spacecraft. By making use of detecting devices called Sun sensors, the mast of Mariner IV was kept pointing at the Sun. The solar-cell panels were always directly facing the Sun, and so were able to collect the maximum amount of solar energy. To keep the spacecraft oriented properly, the sensors, as required, operated jets of nitrogen gas located at the outer end of each of the solar-cell panels. By means of these gas jets, the spacecraft could be rotated about three perpendicular axes to perform

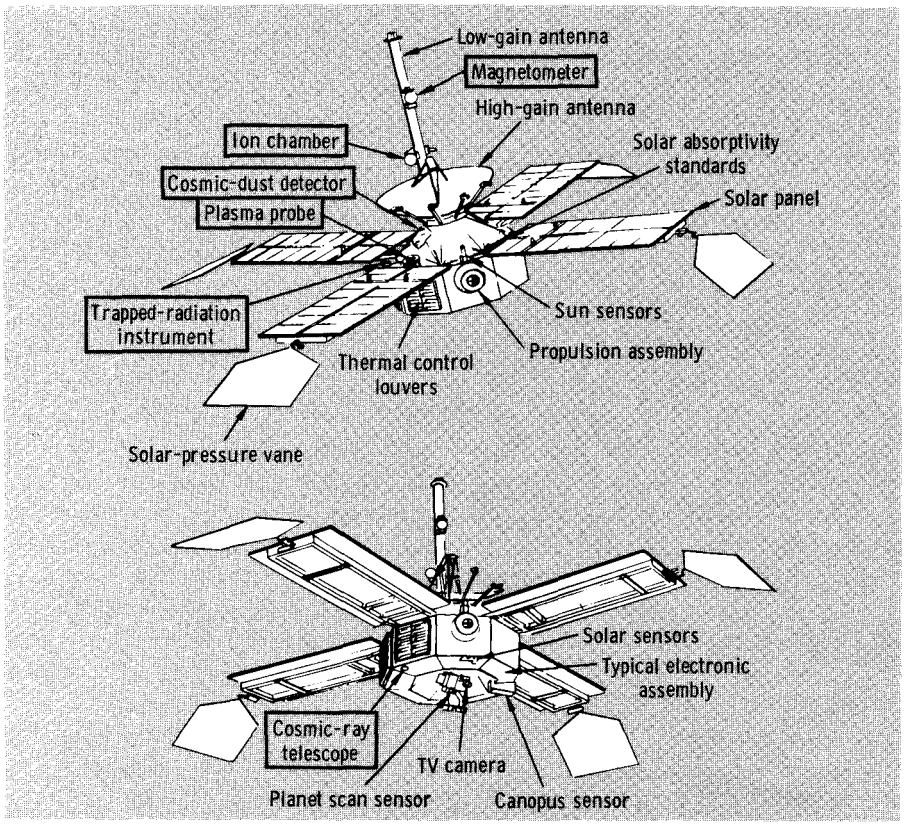


FIGURE 13.7. The Mariner IV spacecraft. (The television camera is at the bottom of the base and is shown in the lower drawing. The titles enclosed in boxes indicate instruments for various scientific experiments.)

roll, yaw, and pitch motions. The purpose of the solar-pressure vanes attached to the ends of the panels was to compensate for the disturbing effect of the pressure of the Sun's radiation and thereby decrease the demand on the gas jets.

Before Mariner IV was launched, the solar-cell panels were folded up and the whole spacecraft was enclosed in a shroud (nose fairing). This served a dual purpose. It protected the spacecraft from damage in its rapid passage through the atmosphere, and it reduced the aerodynamic resistance on it during the launch phase. After launching, the shroud dropped away and when the spacecraft was in its trajectory to Mars, the solar-cell panels opened out and started collecting energy from the Sun. Prior to this, electric power for operating the radio equipment was provided by batteries on the spacecraft. The overall dimensions of Mariner IV, with the panels extended, were 22 feet across and 9½ feet high. The total weight was 575 pounds.

With the Sun sensors keeping the mast (roll axis) directed at the Sun, the spacecraft rotated slowly about this axis. To pre-

vent such rotation and fix the attitude of the Mariner IV in space, it was necessary to make use of another sensor effective in a direction approximately at right angles to that of the Sun. For this purpose, the star Canopus was chosen because of its brightness, which is second only to that of Sirius, and its location close to the south ecliptic pole. With the spacecraft "locked" on the Sun and Canopus, its orientation in space could always be determined with considerable accuracy.

During the first 3 months after launching Mariner IV, radio communication was achieved with the aid of the low-gain, omnidirectional antenna. After this period however, the alinement of the spacecraft with the Sun and Canopus resulted in the high-gain, directional antenna being adequately aimed at Earth (fig. 13.8). Consequently, all radio reception and transmission were performed with the high-gain antenna until several weeks after encounter of the spacecraft with Mars.

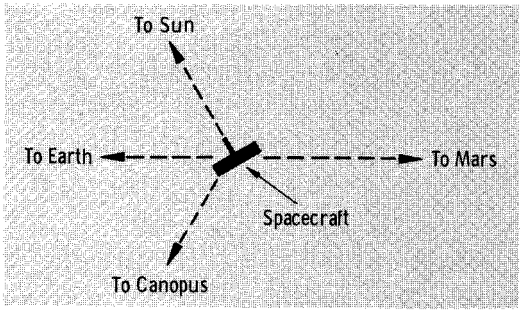


FIGURE 13.8. Orientation of the Mariner IV spacecraft on its way to Mars.

Scientific Instruments on Mariner IV

Some of the instruments on Mariner IV, namely, the solar-plasma probe, the cosmic-ray telescope, and the cosmic-dust detector, were designed for studying the properties of interplanetary space. Others, such as the magnetometer, ionization chamber, and trapped-radiation detector, were intended to provide information about Mars, as well as about space. The magnetometer, which measured the strength of the interplanetary magnetic field, would also have determined the magnetic field of Mars when Mariner IV was in the vicinity of the planet. As already mentioned, the results showed that if Mars has a magnetic field, it must be extremely weak. In agreement with this observation, the ionization chamber and the trapped-radiation detector found no significant radiation belt of electrically charged particles surrounding Mars (p. 64).

As far as Mars itself is concerned, the most important results obtained from Mariner IV were derived from the radio-occultation observations and from the television pictures. The occultation experiment made use of the radio transmitter on the spacecraft and required no special instrumentation. As for the television experiment, it will be described later.

Mariner IV Launch and Trajectory

Mariner IV was launched from the Eastern Test Range, Florida, soon after 14:22 Greenwich mean time (9:22 a.m. eastern standard time) on November 28, 1964, by means of an Atlas-Agena rocket engine combination. The powerful Atlas lower stage carried the Agena with the Mariner IV spacecraft attached to an altitude of about 160 kilometers (100 miles). It was then detached and dropped away. At this point, the nose fairing covering the spacecraft was jettisoned. The Agena engine was ignited for a short time to provide sufficient thrust to place it and Mariner IV into a coasting orbit at the 160-kilometer altitude.

After some 40 minutes in this orbit, when the spacecraft was over the Indian Ocean,

east of South Africa, the Agena was ignited for a second time. The propellants used in the Agena engine are dimethylhydrazine (fuel) and nitric acid (oxidizer), and ignition occurs spontaneously upon mixing. Reignition of the Agena in space, as required by a two-burn trajectory, presented no difficulties. When the desired injection velocity of almost 41 200 km/hr (25 600 mph) was attained, the Agena engine was shut down and detached from the Mariner IV spacecraft. The latter was then in its cruise (unpowered) trajectory to Mars about 45 minutes after the initial liftoff.

During the next few days, the spacecraft was tracked, and from the data obtained by December 4, 1964, it was determined that the trajectory was such as to cause Mariner IV to miss Mars by some 256 000 kilometers (151 000 miles). A calculation was then made of the change in magnitude and direction of the spacecraft speed, at a certain instant, that would be necessary to bring it within roughly 10 000 kilometers (6200 miles) of the surface of Mars. The required midcourse maneuver was then performed in the following manner. The Sun and Canopus sensors were first switched off temporarily so that the attitude of the spacecraft could be changed to the required orientation in space. By utilizing the nitrogen gas jets, Mariner IV was caused to make a pitch turn through an angle of 39.16 degrees and then to roll 156.08 degrees. The midcourse rocket motor in the octagonal part of the spacecraft was then fired for 20.07 seconds.

In this manner, the speed of the rocket was increased by less than 1.5 parts per thousand, and the direction of the trajectory was changed by about ¼ degree. Since Mariner IV still had some 480 million kilometers (300 million miles) to travel before its encounter with Mars, these small changes were sufficient to decrease the miss distance from 256 000 to less than 10 000 kilometers. The rocket motor on the spacecraft had the capability of performing a second midcourse maneuver, but this was deemed to be unnecessary.

The closest approach of Mariner IV to Mars occurred on July 14, 1965, on the 228th day of the flight. The relative locations of Earth, Mars, and the spacecraft on various dates are indicated in figure 13.9. The closest approach distance of Mariner IV to the surface of the planet was 9846 kilometers (6118 miles). The distance from Mars to Earth was then 216 million kilometers (134 million miles). About 78 minutes after the closest approach, the spacecraft passed into the occultation zone behind Mars (fig. 5.6) and remained within it for 45 minutes. During this period radio transmission to Earth was cut off by the planet. Subsequently, Mariner IV continued on its way in orbit around the Sun. Because of the attraction of the gravitational field of Mars, however, this orbit was widened and the period was increased from 529 to 587 days. At the same time, the angle between the plane of Mariner IV's orbit and Earth's orbital (ecliptic) plane was increased from 8 minutes of arc to 2½ degrees.

Mariner IV Television System

The television camera on Mariner IV consisted of a reflecting telescope, with a focal length of about 30 centimeters, which focused an image on the faceplate of a vidicon television tube. The image, only about 5.5 millimeters (0.22 inch) square, was scanned in the usual manner with an electron beam in 200 lines with 200 dots per line. The output of the television camera was recorded on tape and, as will be seen shortly, the information was transmitted to Earth later at a much slower rate. The shutter of the cam-

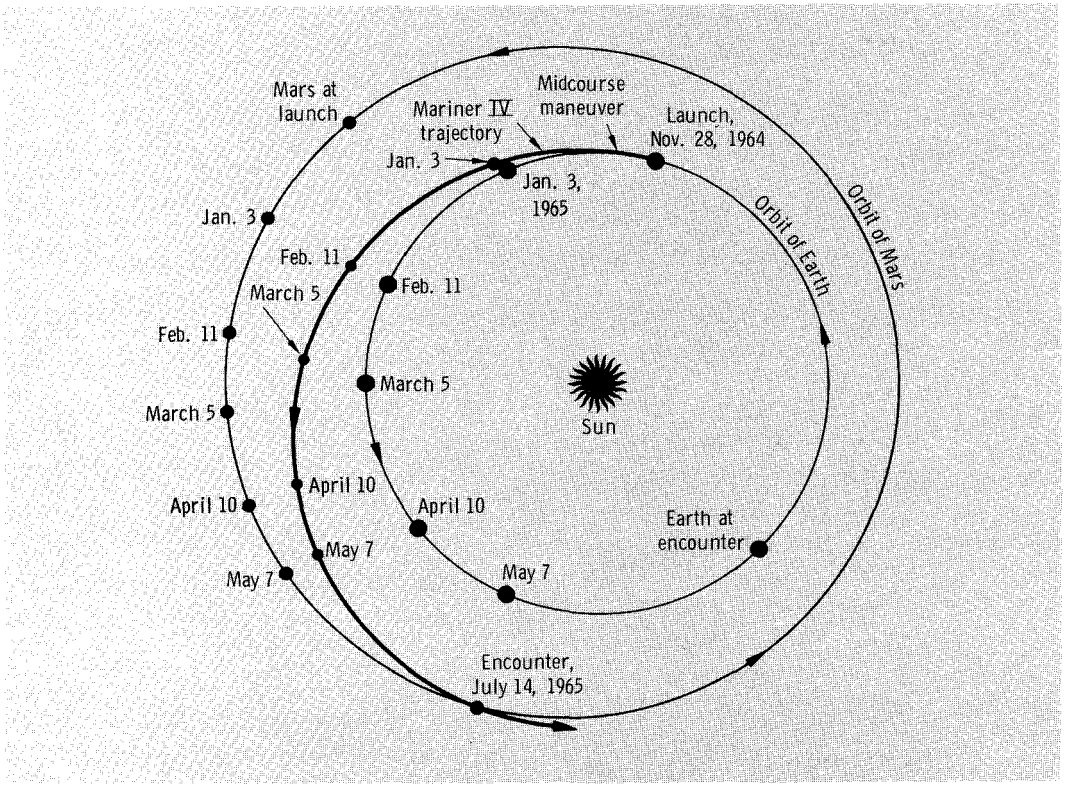


FIGURE 13.9. Relative locations of Earth, Mars, and Mariner IV at various dates.

era was timed to produce an exposure every 48 seconds.

In the hope of obtaining some information about the color of the Martian surface, a disk carrying four light filters, alternately red and green, was rotated through 90 degrees in front of the camera before each exposure. Pictures were not recorded at every exposure, however, because this would have used up the available tape before the scan path of the camera had been completed. To stretch out the picture sequence, while permitting some overlap between adjacent pairs, every third picture was omitted. Thus, the exposure sequence was green-red, red-green, green-red, and so on.

When Mariner IV was getting close to Mars, a wide-angle sensor detecting the sunlight reflected from the planet's surface initiated rotation of the scan platform so as to direct the television camera toward the planet. The platform had actually been set in the approximately correct position by a radio command from Earth in case the automatic sensing system had failed, but it did not do so. A narrow-angle sensor then started the picture-taking procedure, immediately a part of the illuminated disk of Mars came into view. This occurred when Mariner was 16 900 kilometers (10 500 miles) from the surface of the planet. After 22 exposures, in 11 slightly overlapping pairs, were recorded

in about 26 minutes, the television camera was switched off. The spacecraft was then 11 900 kilometers (7400 miles) from Mars.

The areas covered by the 22 photographs from Mariner IV are indicated in figure 13.10. It should be noted that in this illustration north is at the top, although in the pictures of Mars taken from Earth, given earlier in the book, the usual astronomical practice is followed in which north is at the bottom, as seen in a telescope. The actual images on the vidicon tube were square, but the areas observed were not. They varied both in size and shape, because in the early and late pictures the camera viewed the planet at an angle to the vertical. It was only in the middle of the sequence, roughly in the areas marked 9 through 12, that the camera pointed almost directly at the surface.

The first photograph was taken immediately after the narrow-angle sensor detected light reflected from Mars. As will be seen shortly (fig. 13.11), the picture actually shows the limb, that is, the edge of the planet, as it appeared from the spacecraft. The areas covered in the last three pictures—nos. 20, 21, and 22— are seen to be beyond the terminator, where daylight on Mars changed to night (fig. 3.13). These pictures were exposed in the dark and consequently they show nothing.

Because of the great distance from Earth at the time of the encounter of Mariner IV with Mars, it was necessary, in the interest of clear reception, for the data to be transmitted to Earth by radio at a very slow rate. This is the reason the ouput of the television camera was recorded on tape at the time the exposures were made. Soon after the pictures were taken, the spacecraft passed into the occultation region with respect to Earth and it was not until almost 11 hours had elapsed that picture transmission commenced. Mariner IV was then in a favorable position with regard to the receiving antenna at Johannesburg, South Africa. It required 8 hours 40 minutes to transmit each of the 22 photographs. When this was completed, the whole process was repeated to fill in data that may have been lost in the first transmission.

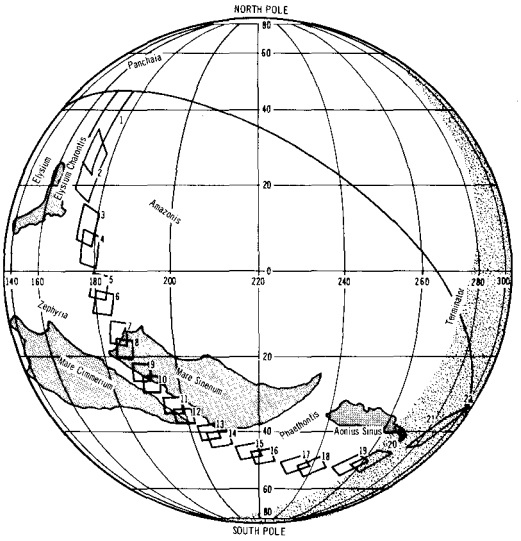


FIGURE 13.10. Areas covered by the photographs from Mariner IV. (North is at the top in this figure.)

Pictures From Mariner IV

The first picture of the surface of Mars taken by the television camera on Mariner IV is reproduced in figure 13.11. North is at the top, as in figure 13.10. It is of historical interest as the first photograph of a planet, other than Earth, ever obtained from a spacecraft. The distance along the limb in the illustration is 660 kilometers (410 miles), and from the limb to the bottom edge of the picture it is about 1290 kilometers (800 miles). The photograph, taken over a bright region, at longitude 165 to 175° and latitude 25 to 45° N, shows relatively light and darkish parts,

FIGURE 13.11. Mariner IV Picture No. 1; north is at the top.

FIGURE 13.12. Mariner IV Picture No. 3; north is at the top.

but there is no sign of the Hades canal which crosses the area.

A peculiar feature of the first picture, as received, is that it shows what appears to be a haze, extending to an altitude of over 100 kilometers (62 miles), in the black sky just above the limb at the right (east) end. It is not known if this is due to a fault in the picture-transmission system or if it is something in the Martian atmosphere. One possibility is that the haze may represent a cloud of small particles of solid carbon dioxide suspended in the atmosphere.

The second Mariner IV picture shows light and dark areas, but it is not until the third (fig. 13.12) that surface features become apparent. The area depicted is between 170° and 180° longitude and 5 and 15° N latitude, and measures 350 kilometers (220 miles) from east to west and 500 kilometers (310 miles) from north to south. In the photograph there are seen a few craters of moderate size and some dark rilles (from the German, *rille,* meaning furrow or groove) which resemble those on the Moon.

The next three pictures, taken over the bright Zephyria region, show only moderate surface detail. A contributory factor may have been that, during the early exposures, the Sun was fairly high in the Martian sky and hence there were no marked shadows. Under such conditions, as is well known from lunar photographs, surface features are difficult to identify.

Pictures 7 through 14, some of which were given in chapter VI, are the most interesting. They cover both bright and dark areas of Mars and show numerous craters ranging in diameter from 175 down to 2.8 kilometers (110 to 1.7 miles). There is some indication of part of the wall of a possible crater with a diameter of 350 kilometers (217 miles) in picture 13, and there are undoubtedly many

craters smaller than 2.8 kilometers across which cannot be resolved in the photographs.

In addition to the craters and rilles, some of the Mariner IV photographs show straight features called lineaments ranging from about 160 to 320 kilometers (100 to 200 miles) in length. As stated in chapter VI, these cannot be identified with the canals, but there are indications of dark, irregular features which might possibly be parts of the canals.

The Mariner IV photography covered less than one-hundredth part of the surface of Mars, and this may be one reason there was no clear indication of the canals. It should be noted, however, that the pictures do not show any striking differences between established bright and dark areas. Furthermore, there is not much difference between exposures of overlapping regions made through red and green filters. The conclusion drawn is that most of the surface of Mars, bright and dark, is reddish in color, although the dark areas are somewhat more red than the bright ones.

THE MARINER 1969 MISSIONS

Description of Spacecraft

The proposed Mariner mission to Mars, planned for the opposition period in 1969, is referred to as Mariner Mars 1969. Like that of Mariner IV, the mission involves a flyby trajectory, but with a closer approach to the surface of Mars, possibly to within about 3000 kilometers (1850 miles). The basic design of the spacecraft (fig. 13.13) is similar to that of Mariner IV. It consists of an octagonal structure, containing the electronic equipment and the rocket motor for the midcourse maneuver. There are four solar-cell panels, but the solar-pressure vanes have been eliminated. Sensors for detecting the Sun and the star Canopus provide orientation of the vehicle in space. The mast is kept pointing toward the Sun so that the panels receive maximum exposure to solar radiation, as in Mariner IV. There are also, as before, an omnidirectional low-gain antenna on the mast and a directional, high-gain antenna on the octagonal structure. With the solar-cell panels fully extended, the overall dimensions are 19 feet across and 11 feet high, and the total weight of the Mariner Mars 1969 is almost 840 pounds.

As far as instrumentation is concerned, there are three main differences between the Mariner IV and Mariner 1969 spacecraft. First, all the instruments in the latter, which are mounted on a scan platform underneath the octagonal structure as viewed in figure 13.13, are intended exclusively for the exploration of Mars and not to study interplanetary space. Second, there are two television cameras on the Mariner Mars 1969. One covers a larger area but shows less detail than the other. Finally, the scan platform carrying the instruments and the television cameras (fig. 13.14) is capable of rotation in two directions at right angles, whereas in Mariner IV the movement was restricted to one direction.

Other important advances in the Mariner Mars 1969 spacecraft are concerned with the storage and transmission to Earth of data obtained by the television cameras and scientific instruments. The data-storage capacity has been increased by a factor of about 35 over that of Mariner IV. One consequence will be that more and better television pictures can be obtained of the Martian surface. Furthermore, the rate of transmission of data to Earth has been increased greatly. This has been made possible by increasing the power of the radio transmitter and the dimensions of the high-gain antenna on the spacecraft, by using the 210-foot antenna of the Deep Space Network at Goldstone, Calif.,

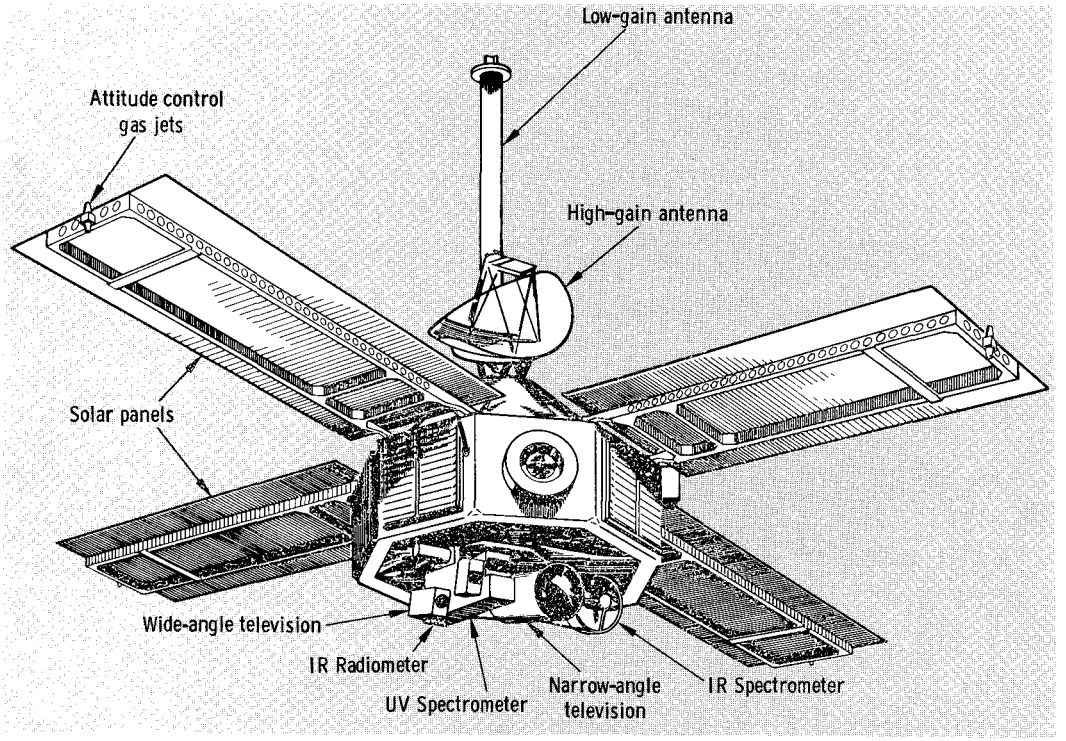


FIGURE 13.13. The Mariner Mars 1969 spacecraft.

completed in 1966, for receiving the signals, and by various design changes.

In addition to the two television cameras, the scientific instrumentation on the Mariner Mars 1969 consists of an infrared spectrometer, an infrared radiometer, and an ultraviolet spectrometer, as shown in figure 13.14. The types of observations that can be made with these instruments were outlined earlier in this chapter. Some details concerning the actual devices to be carried by the Mariner Mars 1969 spacecraft are given below.

Television System

The television system on the 1969 Mariner spacecraft consists of two vidicon-tube cameras, one of 5 centimeters focal length having a wide-angle field of view with a low resolution, and the other with a focal length of 50.8 centimeters being a narrow-angle device giving a high resolution. The latter views a much smaller area, but greater detail will be visible in the pictures obtained. Red, blue, and green filters will be placed alternately in front of the wide-angle camera before exposure, but because of the greater capacity for storage, it will not be necessary to omit any of the pictures, as was the case for Mariner IV.

Television pictures will be taken in two phases, called far-encounter and near-encounter, respectively. The far-encounter phase will commence 2 or 3 days before the spacecraft reaches the point of closest approach to Mars. The distance to Mars at the beginning of this phase will be roughly 1.6

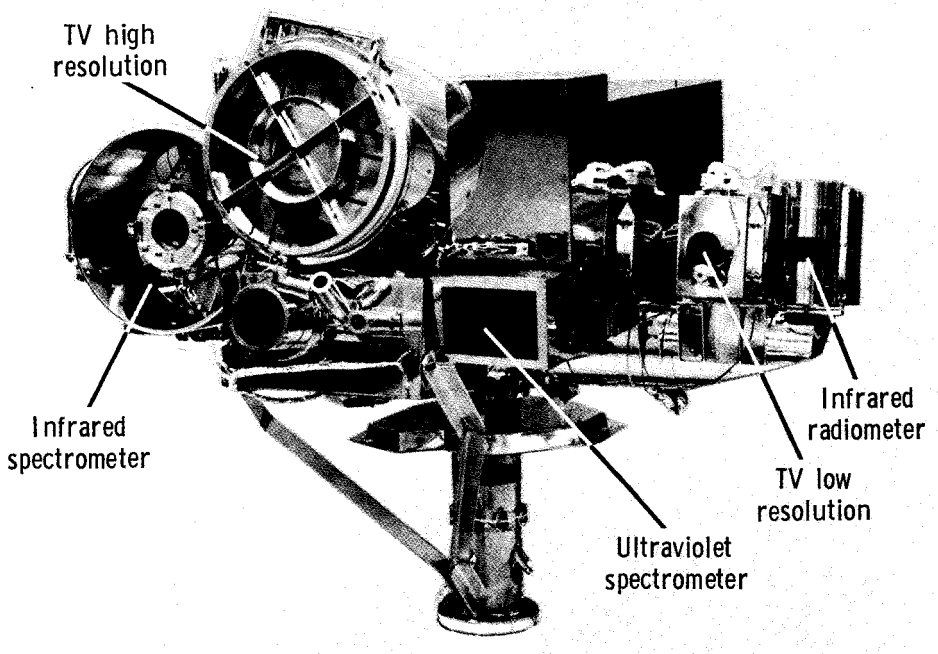


FIGURE 13.14. Scan platform and instruments of the Mariner Mars 1969 spacecraft.

million kilometers (1 million miles). The scan platform will then be moved into a position where a sensor can detect sunlight reflected from the planet and keep the high-resolution camera pointed in the required direction. A series of television pictures will be taken, one every ½ hour. They will show the full disk of Mars, but with a much greater resolution than can be achieved from a telescope on Earth. As the planet rotates, almost all the surface will come into view of the television cameras while pictures are being taken. The later pictures in the far-encounter phase will show more detail than the earlier ones, because of the much smaller distance from the spacecraft to Mars.

The near-encounter phase will start within about half an hour of the closest approach to the planet. Just before the commencement of this phase, the scan platform will be moved into such a position as to provide the desired scan path for the cameras and instruments across the face of Mars. Both high- and low-resolution pictures will be taken during the near-encounter phase. The total number of exposures in this phase is expected to be about 25, with some overlap between wide angle pairs.

When the planet, illuminated by the Sun, comes into view of the cameras, a sensor on the spacecraft will start the process of picture taking. After about 6 minutes, the scan platform will be rotated to permit a more southerly area of the planet, possibly includ-

ing the south polar cap, to be observed (fig. 13.15). This type of rotation was not feasible in Mariner IV. Before the scan path reaches the terminator and enters the local nighttime region of Mars, the platform will be placed in a new position which will permit a few more exposures to be made of the sunlit portion of the planet.

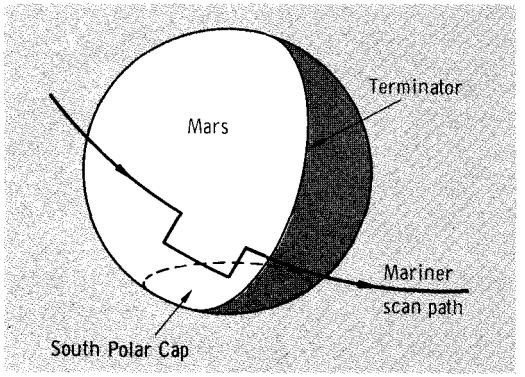


FIGURE 13.15. Proposed scan path of one of the Mariner Mars 1969 spacecraft. (North is at the top in this figure.)

Scientific Instruments

In the infrared spectrometer on Mariner Mars 1969, radiation reflected by and emitted from the surface of Mars will be collected by a telescope of 25-centimeter focal length. The infrared radiation will then be divided by means of a filter-wedge arrangement into two parts, with wavelengths in the ranges of 1.5 to 6.0 μ and 4.0 to 15 μ, respectively. Absorption bands of organic compounds, such as that of the C—H bond, for example, and of carbon dioxide occur in the former range, whereas bands of water vapor, methane, ethylene, and ozone are found in the latter range of wavelengths.

The absorption spectrum is obtained in each range by measuring the infrared radiation intensities at a series of wavelengths. For the range from 1.5 to 6.0 μ, the infrared detector is lead selenide (PbSe) maintained at a temperature of 130° K, that is, −143° C, by radiating heat to space. In the wavelength range of 4.0 to 15 μ, the detector is made of mercury germanide (HgGe). This is cooled to 27° K (−196° C) in two stages by utilizing the Joule-Thomson (throttled expansion) principle with nitrogen and hydrogen gases in turn.

The general scan path of the infrared spectrometer will be the same as that of the television cameras during the near-encounter phase of the Mariner Mars 1969 mission. It will thus be possible to correlate any unusual features in the infrared spectrum, such as indications of abnormal amounts of carbon dioxide or water vapor in the atmosphere, with specific regions on the surface of the planet. The spectrometer is designed to take a complete spectrum in each wavelength range every 10 seconds. Continuous operation will thus yield almost 200 infrared spectra from 1.5- to 15-μ wavelength. The television cameras will be switched off when the scan path crosses the terminator and enters the nighttime region of Mars. But the infrared spectrometer will continue to operate, since infrared radiation is emitted by the dark areas of the planet.

An infrared radiometer on the 1969 Mariner spacecraft will measure the infrared radiation emission from Mars in two wavelength ranges for the purpose of determining surface (or cloud) temperatures. The two ranges are 8 to 12 μ and 18 to 25 μ, thereby avoiding regions in which significant atmospheric absorption might occur. The amount of energy emitted in each wavelength range will be measured with a bismuth-antimony (Bi-Sb) thermopile. From the results, the temperatures of the emitting surfaces will be calculated.

The scan path of the infrared radiometer will be the same as that of the televison cameras and the infrared spectrometer. It will thus be possible to relate surface features with their temperature, as well as with the carbon dioxide, water vapor, and possibly other gases in the local atmosphere.

Two aspects of the infrared radiometer measurements are of special interest. First, during the southerly scan by one of the two Mariner Mars 1969 spacecraft, it is hoped to bring the south polar cap under observation. If this can be done, infrared thermal emission may make it possible to determine whether the main constituent of the cap is solid carbon dioxide or hoarfrost (solid water). Second, the infrared measurements will be continued beyond the terminator in order to observe the surface temperature of Mars at various times after local sunset. This is something that cannot be done with any degree of certainty from Earth.

The ultraviolet spectrometer for the 1969 mission to Mars will consist of a telescope with a focal length of 25 centimeters combined with an Ebert-type spectrometer having a moving grating. The instrument has two channels covering the wavelength range from 1100 to 4300 Å; this extends into the visible (violet) part of the spectrum and includes all the lines in the tabulation on page 281. The radiation intensities at individual wavelengths are determined by means of photomultiplier tubes. The ultraviolet spectrometer will be pointed in such a way that regions of different altitude in the Martian atmosphere will come into view as the spacecraft flies by the planet.

An occultation experiment will be performed in connection with the radio transmission from the Mariner 1969 spacecraft as it passes into the zone where Mars cuts off the radio signals. This requires no special equipment on the spacecraft and is similar to the occultation experiment performed on Mariner IV. The data may be better than in the latter case because of the shorter distance between Mars and Earth at the time of encounter and the improvement in the radio-communication system between the spacecraft and Earth.

Launch and Trajectories

It is proposed that two Mariner spacecraft be launched to Mars in 1969, at an interval of 4 or 5 weeks. The spacecraft are identical, but the trajectories, which are both of type I, will bring different parts of the planet under observation. The launch vehicle is to be the Atlas-Centaur combination. The Centaur upper stage is more powerful than the Agena engine used for the Mariner IV launch, and so an increased payload is possible. The Centaur actually consists of two rocket engines with liquid hydrogen as the fuel and liquid oxygen as the oxidizer. Unlike the propellants in the Agena engine, those in the Centaur do not ignite when they are mixed. A special ignition system must therefore be used.

Reignition of the Centaur engine in space, as would be necessary for a two-burn mode of injection, involves some uncertainty. In addition, tracking of the trajectory near Earth will be more reliable in the direct ascent (one-burn) mode. Consequently, it has been decided to use a single burn of the Centaur in launching the Mariner Mars 1969 spacecraft. Thus, there will be no coasting orbit with the upper stage engine not operating, as was the case with Mariner IV. The launch time will be selected to allow direct injection into the required Martian trajectory.

The Atlas lower stage will be essentially the same as in the Atlas-Agena combination. It will be detached and will drop away completely after about 4 minutes of operation. The nose cone protecting the spacecraft will

then be jettisoned and the Centaur engines started. The Centaur will burn continuously for a few minutes, after which it will separate from the Mariner spacecraft. The latter will then be on a cruise trajectory designed to bring it near Mars.

About a week after launch, when the trajectory has been established by repeated tracking, a midcourse maneuver will be performed. If it proves necessary, a second change in the trajectory will be made later to bring the Mariner spacecraft within about 3000 kilometers (1850 miles) of the surface of Mars at closest approach. After passing by the planet, the trajectory will curve around the far side with reference to Earth to permit occultation of the radio signal from the spacecraft.

The launch period for type I trajectories to Mars of the Atlas-Centaur-Mariner combination during the 1969 opportunity extends from about February 16, 1969, to April 9, 1969. During the early part of this period, the initial portion of the trajectory, near Earth, is so low over the Southern Hemisphere that it cannot be tracked some of the time from the stations of the Deep Space Network, but the tracking situation will improve significantly during the course of the launch period.

The first Mariner Mars 1969 is scheduled to be launched from the Eastern Test Range in Florida late in February 1969. For the trajectory which is to be used, encounter with Mars should occur during the first week of August 1969. The flight time is to be relatively short, roughly 150 to 160 days. The actual arrival date will be selected prior to launch and the trajectory adjusted accordingly. The television and instrument scan path planned for the first spacecraft will be along the equatorial region of Mars during the middle of the local spring. It is on this mission that rotation of the scan platform in a southerly direction (fig. 13.15) should make it possible for the television cameras and the scientific instruments to view the south polar cap.

The second Mariner Mars launch is expected to take place about a month or so after the first, but necessarily before April 9, 1969, which marks the close of the launch period. The trajectory will be shorter than that of the Mariner launched earlier in the year, and encounter should occur around August 5, 1969. The area scanned by the spacecraft is expected to be mainly in the southern region of Mars. Consequently, if both spacecraft are launched and operate successfully, they should provide information about several parts of the Martian surface that are of special interest from both physical and biological points of view.

FOR FURTHER INFORMATION

BARRINGTON, A. E.; CARUSO, A. J.; AND MANNELLA, G. G.: Flyby Measurements of the Profiles of Water Vapor, Particulate Matter, and Carbon Dioxide in the Martian Atmosphere. NASA TN D–4115, 1967.

GEORGIEV, S.: A Feasibility Study of an Experiment for Determining the Properties of the Mars Atmosphere. Vol. I, Summary, NASA CR–530, 1967.

JAMES, J. N.: The Voyage of Mariner IV. Sci. Am., vol. 214, no. 3, 1966, p. 42.

LEIGHTON, R. B.: The Photographs From Mariner IV. Sci. Am., vol. 214, no. 4, 1966, p. 54.

LEIGHTON, R. B.; ET AL.: Mariner Mars 1964 Project Report: Television Experiment, Part I: Investigators' Report. Jet Propulsion Laboratory Tech. Rept. no. 32–844, Pt. I, 1967.

Mariner IV Mission to Mars. Jet Propulsion Laboratory Tech. Rept. no. 32–762, Pts. I and II, 1965. (Reprinted from Astronaut. Aeron., vol. 3, nos. 8 and 10, 1965.)

Mariner IV Measurements Near Mars: Initial Results. Jet Propulsion Laboratory Tech. Rept. no. 32–833, 1965. (Reprinted from Science, vol. 149, 1965, pp. 1226–1268.)

NICKS, O. W.: A Review of the Mariner IV Results. NASA SP–130, 1967.

PICKERING, W. H.: Navigating the Mariner Spacecraft to Mars. Proc. Am. Phil. Soc., vol. 110, 1966, p. 332.

Report from Mars: Mariner IV 1964–1965. Prepared by Jet Propulsion Laboratory, U.S. Government Printing Office, 1966.

SEIFF, A.; AND REESE, D. E.: Defining Mars' Atmosphere—A Goal for Early Missions. Astronaut. Aeron., vol. 3, no. 2, 1965, p. 16.

SOMMER, S. C.; ET AL.: The Structure of an Atmosphere From On-Board Measurements of Pressure, Temperature, and Acceleration. NASA TN D–3933, 1967.

SOMMER, S. C.; AND BOISSEVAIN, A. G.: Atmosphere Definition With a Free-Falling Probe. Astronaut. Aeron., vol. 5, no. 2, 1967, p. 50.

Index

A

Absorption spectra, 73–74, 79
Acceleration, gravitational. *See* Gravity, of Mars.
Accelerometer, 278
Acetaldehyde, 181
  in Martian atmosphere, 81
Acidic group. *See* Carboxyl group.
Acrylonitrile, 186
Adenine, 165 ff.
  stability, 187
  synthesis, laboratory, 187
Adenosine diphosphate, 189
Adenosine monophosphate, 189
Adenosine triphosphate, 189–190
  for life detection, 235
Adenylic acid, 187
Adiabatic lapse rate, 95–96
ADP. *See* Adenosine diphosphate.
Aerodynamic resistance, 278
Aerosol, 227
Agena, 287 ff.
Aiming point, 277
Alanine, 163, 173, 180, 184
Albedo, 64
  bright areas, 109
  dark areas, 111–112
  Mars, 65–66
  radiometric (or integral), 65
  snow, 104
Aldehydes, formation, 180
Almagest, 9
Alpha amino acids, 163
Alpha-particle scattering, 245, 284
Altitude
  pressure, barometric, variation, 91–92
  surface. *See* Elevation.
Amide, 181
Amino acetamide, 181
Amino acid amide, 181
Amino acids, 162
  conjugated, 163–164
  formation, 179–182
  number, 162–163
  optical activity, 164–165
  residue, 164
  simplest, 162
  synthesis, 165
Amino group, 162
  alkalinity, 163
Amino nitriles, 181, 183
Amino propionamide, 186
Aminoacetic acid, 162
Aminoacetonitrile, 184
Ammeline, 201
Ammonia, 74
  Martian atmosphere, 81
  in primitive atmosphere, 178
Ammonium cyanate, 186
Ammonium cyanide, 182
AMP. *See* Adenosine monophosphate.
Anaerobe, 178, 213–215
Angle of inclination, 17, 47
Angle, phase, 43–44
Angstrom unit, definition, 79
Angular diameter, 57–58
Anticodon, 173
Antigen reaction, 185

Aphelion (*see also* Conjunction; Opposition; Perihelion), 46, 54
brightness, effect on, 41
definition, 36
distance, 36
Mars, 36
speed, 36
temperature, 132
Apparition, 21
Archebiosis, 157
Areography, 8, 19
Areocentric, 8
Ares, 8
Argon
carrier gas, 229
Martian atmosphere, 78, 82, 89
origin, of atmospheric, 75, 82
quantity, of atmospheric, 75
Ascent, direct, 277
Aspartic acid, 180, 184
Asteroid, 33–34
belt. *See* Asteroidal belt.
discovery, 34
eccentricity, orbital, 35
formation, 34
perturbation of, 57
size, 33
Asteroidal belt, 33, 111
carbonaceous chondrites, 205, 210
craters, of Mars, 128
meteorites, source, 111
Astronomical unit, 58
Asymmetric synthesis, 195
Asymmetry, molecular (*see also* Optical activity), 164
amino acids, 164–165
origin, 195
ribose, 166
Atlas, 287 ff.
Atmosphere, of Earth
origin, 75–76
ozone, 95
primitive, 177–179
reducing, 183
Atmosphere, of Mars, 73–97
acetaldehyde, 81
ammonia, 81
argon, 81, 89
blue haze. *See* Haze, of Mars.
carbon dioxide, 76–78
carbon dioxide, crystals, 145
carbon dioxide, mass, 105
carbon monoxide, 81
clouds. *See* Clouds, of Mars.
composition, 81
density, 91
formaldehyde, 81
gases, escape, 210
gases, identification, 73–74
gases, miscellaneous, 81–82
haze. *See* Haze, of Mars.
hydrogen sulfide, 81
ionosphere, 92–94
ions, 94, 148–149
methane, 81
nitrogen dioxide, 81, 109, 149
oxygen, 81–82
ozone, 81–82
particles, 144–148
pressure. *See* Pressure, barometric.
pressure variation, 91–92
primitive, 210
scale height, 89
solid particles, 86
temperatures, 89, 94–97
water vapor, 78–81
Atmosphere, planetary, 73–76
gases, identification, 73–74
primitive, 177–179
Atmosphere, reducing. *See* Reducing atmosphere.
Atmosphere, standard, 77
Atmospheric density, 91
Atmospheric pressure. *See* Pressure, barometric.
Atmospheric probe, 273, 278–280
ATP. *See* Adenosine triphosphate.
AU. *See* Astronomical unit.
Autocatalysis, 159
Automated biological laboratory, 242–246
atmosphere, observations, 242–244
biological experiments, 245
concept, 242
soil, studies, 244–245
Autotrophic organisms, 214
Autumn, 51
Martian, 53–55
Autumnal equinox, 51
Axis
Earth, 47
ellipse, 35
inclination, 17, 47, 49–55
major orbit, of Mars, 35, 55

B

Bacteria, 159
  anaerobic, 178
  chemosynthetic, 215
  coccoid, 193
  encapsulated, 216
  halophilic, 213, 218
  halophytic, 213, 218
  photosynthesis, 214
Bar, definition, 76–77
Barometric pressure. *See* Pressure, barometric.
Base, nitrogenous, 165
Beneficial mutation, 159
Beta-alanine, 186
Beta-amino propionamide, 186
Beta-amino propionitrile, 186
Bible, origin of life, 154
Biological exploration, 225–229
Biological laboratory, automated. *See* Automated biological laboratory.
Biology, molecular, 161–176
Blackbody, 131 ff.
Blue clearing (*see also* Haze, of Mars), 143
  carbon particles, 147
  ion recombination, 148–149
  nitrogen dioxide, 149
  opposition, 149, 151
  protons, high-energy, 149
  rapidity, 147
Blue haze. *See* Haze, of Mars.
Bode's rule, 33–34
Boltzmann gas constant, 89
Bond albedo, 65
Bond, peptide, 164
Bright areas (*see also* Dark areas), 109–111, 121–125
  albedo, 65, 109
  cloud location, 140
  color, 109–110
  crater density, 128
  dark area, inside, 116
  duststorms, 141
  elevation, 90, 121–125
  infrared reflection, 110
  iron, 111
  Mariner IV photographs, 292
  meteorites, 111
  reflectivity, 112
  rock, 110–111
  silicate, 111
  temperature, 121
Brightening, limb, 58–59
Brightness. *See* Albedo.
Bulge, equatorial, 62

C

Calcite, 85
Calcium hydroxide, 188
Calorimetry, for life detection, 240–241
Canale, 22
Canals, 21–29, 125–128
  cloud location, 140
  "gemination," 24–25, 29
  radar measurements, 127
  seasonal change, 29, 127
  seeing conditions, 125–126
  "twinning," 24–25, 29
  wave of darkening, 127
Carbohydrates
  fermentation, 214
  oxidation, 214
Carbon dioxide, 89
  crystals, in atmosphere, 145
  decomposition, radiation, 81
  density, of solid, 106
  Earth's atmosphere, 74
  Martian atmosphere, 76–78, 82, 89
  mass, in atmosphere, 105
  partial pressure, 106, 108
  polar caps, 103–109
  solid, density, 106
  sublimation, 106
Carbon-14 labeling, 236
Carbon monoxide, in Martian atmosphere, 81
Carbonaceous chondrites, 196–206
  carbon compounds, 200–202
  composition, 197
  conclusions, 205–206
  contamination, 200
  elements, organized, 203–205
  extraterrestrial life, relevance to, 198
  fossils, 156, 203
  hydrocarbons, 198–200
  optical activity, 202–203
Carbonates, Martian surface, 220
Carboxyl group, 162
  acidity, 163
  fatty acids, 202
Carrier gas in chromatography, 229
Cassini's drawing, 17
Catalysis, 159
  clay, 184
  condensing agent, 185
  enzyme, 162
  solid surface, 175

Cell, 159–161
division, simulation, 193
structure, 160
wall, 160
Centaur, 296, 297
Chaldeans, 8
Charged particles, 92–94
near Mars, 64
Chemosynthetic bacteria, 215
Chlorin, 201
Chlorophyll, 200
Chondrites, 111
carbonaceous. *See* Carbonaceous chondrites.
Chondrules, 197
Chromatin, 160
Chromatography, 283
gas. *See* Gas chromatography.
Chromosome, 160–161, 168, 171
Chymotrypsin, 185
Cirrus clouds, 137
Clairaut ratio, 66, 68
Clay, catalysis, 175, 184, 188
Clearing, blue. *See* Blue clearing.
Clouds (*see also* Clouds, of Mars), 18
cirrus, 137
condensation, 122
formation, 121–122
noctilucent, 148
Clouds, of Mars (*see also* Haze, of Mars), 137–143
color, 137
composition, 141
dimensions, 140
duration, 141
duststorms, 141–142
formation, 139
localized, 139–140
morning, 139
movement, 140–141
nitrogen dioxide, 143
opacity, 141
polar hood, 139
recurrent, 140
size, 140
stationary, 139
types, 137–138
W-cloud group, 140
white, 138–141
yellow, 141–143
Coacervates, 191–194
Coasting orbit, 277
Coccoid bacteria, 193
Codon, 172–173
Colloidal solution, 192
Colonization, of Mars, 220–223
food, 222–223
habitation, requirements, 221–222
power, 222–223
Color, of Mars (*see also* Albedo), 83
bright areas, 109–110
dark areas, 112–113
Comet, 205
debris, 147
impact on Mars, 82
Committee on Space Research. *See* COSPAR specifications.
Condensation, 188
Condensing agent, 185
Confidence, of noncontamination, 252
Conjugated amino acid, 163–164
Conjunction (*see also* Aphelion; Opposition; Perihelion), 37 ff.
definition, 37
phases, relation to, 43
Constellations, 7–8
Contamination, of Mars, 218
avoidance, 247–271
constraints, relaxation of, 255–257
COSPAR specifications, 247–258
model, 257–258
modes, 248–250
risk, 250–254
sources, 250–255
Contamination, of meteorite, 200
Continents, 21, 109
Contrast, surface. *See* Surface-contrast theory.
Copernican planetary system, 10–12, 15
Core, of Earth. *See* Interior, of Earth.
Core, of Mars. *See* Interior, of Mars.
COSPAR quarantine. *See* COSPAR specifications.
COSPAR specifications, 247–258
basis, 250–254
decontamination, 258–259
implementation, 258–271
sterilization, 258–259
Cosmic-dust detector, 287
Cosmic-ray telescope, 287
Cosmic rays, 179, 215–217
primary, 244
Cracking pattern, mass spectrum, 232
Craters, 128–131
age, 131, 211
density, 128
diameter, 128, 130

meteoritic origin, 131
number, 130
polygonal, 130
vulcanism, 130
Crust, 62, 67
Cyanamide, 184
polymerization, 201
Cyanoacetylene, 186
Cysteine, 164
Cytoplasm, 160
Cytosine, 165 ff.
carbonaceous chondrites, 198
prebiotic synthesis, 186
uracil, conversion to, 186

D

Dark areas (*see also* Bright areas), 111–125
albedo, 65, 111–112
boundaries, 115
bright area, within, 116
changes, 113–117
cloud location, 140
clouds, avoidance by, 123
color, 112–113, 120
crater density, 128
darkening, wave of. *See* Wave of darkening.
duststorm, 118
elevation, 90, 121–125
Mariner IV photographs, 292
nature of, 120–121
particle size, 120
radar reflection, 123–124
reflectivity, 112–113, 120
reflectivity, factors affecting, 118
refraction, index of, 121
seasonal change, 112–113
secular change, 113, 116
silicates, 120
structural details, 115–116
temperature, 121
vegetation, 111
wave of darkening. *See* Wave of darkening.
Darkening, limb, 58–59
Darkening, wave of. *See* Wave of darkening.
Dawes map, 19–20
Day
duration, 51
Mars, 16, 63–64, 72
names, 8
sidereal, 63
solar, 63
solar, mean, 64
Decontamination, of spacecraft, 258–259
Deferent, 9–10
Dehydration, of minerals, 75–76
Deimos. *See* Moons, of Mars.
Deleterious mutation, 159
De Mottoni's map, 28, 31
Density
atmospheric, 91
of Mars, 62–63, 72
number, 89, 91
Deoxynucleotide, 166
Deoxyribonucleic acid, 168–171
configuration, 169
magnesium ions, 175
mitosis, 169
mutation, 170–171
polymerase, 174
primer, 174
radiation damage, 211
replication, 170
structure, 169
synthesis, 174
Deoxyribose, 165
Depth, effective, atmospheric, 77
Deserts, 109
Dextrorotation
definition, 164–165
sugars, 166
Diameter, of Mars
angular, 57–58
apparent, 41
determination, 57–59
equatorial, 59–60
linear, 58
polar, 59–60
Diastereoisomers (*see also* Optical activity), 233
separation, 239
Dicyanodiamide, 184
Dieaway, of micro-organisms, 266
Differentiation, in interior, 67–68, 82
Dimer, 184
Diogenes, 235
Dipeptide, 163
Direct ascent, 277
Direct motion, 44–46
Disturbance. *See* Perturbation, by Mars.
DNA. *See* Deoxyribonucleic acid.
Doppler effect
definition, 79
water vapor detection, 79
wavelength shift calculation, 79
Doppler shift. *See* Doppler effect.

Dose rate, 215
Drag, aerodynamic, 278
Dust
  color, 118
  particles, 59
  saltation, 119
Duststorm, 17, 141
  height, 142
  hemisphere, 142
  perihelion, 141
  recovery from, 219
  wave of darkening, 118
Dye spectra, shift, 234
Dynamical flattening. *See* Flattening, dynamical.
Dyne, definition, 76

E

Earth
  atmospheric temperatures, 95
  interior, 65, 67
  Mars, flight time to, 276
  mass, 57
  orbit, 14
  polar caps, 99, 104
  rotation, 10
  seasons, 50–52
Eastern quadrature, 41–42
Eccentricity, 35
  definition, 35
  Mars' orbit, 35
  opposition, effect on, 39
Eclipse, 46, 71–72
Ecliptic, 46
Ecliptic plane, 46
Effective depth, atmospheric, 77
Electron acceptor, 213
Electron density, 92
Electron-density scale height, 96–97
Electron donor, 214
Electron microprobe, 205
Electron microscope, 241–242
Elevation
  bright and dark areas, 121–125
  differences, on Mars, 90
  general, for Mars, 124
Ellipse
  axis, 35
  circle, relation to, 35
  definition, 34–35
  eccentricity, 35
  foci, 34–35
  planetary orbit, 13–15
  radius vector, 14
Elliptical orbit (*see also* Ellipse), 34–36
Ellipticity. *See* Flattening, dynamical.
Elongation, 42
Emitter, ideal. *See* Blackbody.
Energy, nuclear, 223
Enzyme, 162, 174–175
  hydrolytic, 193
  proteolytic, 185
Epicycle, 9–10
Equator
  centrifugal force, 66
  position of, 19
Equatorial bulge, 62, 67
Equatorial diameter, 59–60
Equatorial plane, inclination, 47
Equilibrium, temperature, 131
Equinox, 51–53
  polar caps, 99–100
Equipotential surface of gravity, 90–91
Erosion
  craters, destruction of, 130
  wind, 118, 119
Escape velocity, 277
Escherichia coli, 215
Ether, diethyl, 185
Ethyl metaphosphate, condensing agent, 185, 190–191
Evaluator, multiple. *See* Multivator.
Evening star, 43
Exobiology, of Mars, 209–212
  definition, 209
Exosphere, 95–96
  temperature, 210

F

Fall, meteorite, 196
Fatty acids
  in carbonaceous chondrites, 201
  in nature, 202
Faults, on Mars, 127
Features, of Mars
  naming, 20–21
  table, 26–27
Fermentation, 213
Ferric oxide
  bright areas, 109–111
  dark areas, 120–121
  hydrated, 110
  oxidizing agent, 213
Filter, light, 84
  clouds, observation, 137
  colored (*see also* Wright effect), 58–59
  haze, study, 143

Fireball, 196
Fisher-Tropsch process, 200
Fission, binary, 160
  deoxyribonucleic acid, 169
Fixed stars, 7, 10
Flammarion's map, 21
Flares, solar, 215
Flattening, dynamical, 60–62, 66, 72
Flattening, of Mars. *See* Flattening, dynamical.
Fluorescence, 216
Fluorescence, for life detection, 238
Flyby vehicles, observations, 280–282
Flying spot, 241
Foci, of ellipse, 34–35
Fontana's drawings, 15–16
Formaldehyde, 181
  hydroxylamine, reaction with, 183
  Martian atmosphere, 81
  sugar formation, 188
Formamide, 182
Formate, labeled, 236
Forward scattering, 145–146
Fossils
  age, 176
  in meteorites, 156
Fractionation, of surface particles, 122
Fractures, in Martian crust, 127
Frost, 78
  craters, 128
  deposition, 121
  polar caps, 103–105
Fungi, 204
  ultraviolet radiation, 216

G

Gas chromatography, 229–231
  detectors, 230
  diastereoisomers, separation, 239
  mass spectrometry, combination, 232–233
Gas constant
  Boltzmann, 89
  universal, 89
Gehon, 17
Geiger counter, 236
"Gemination," canals, 24–25
Gene, 160
Generation, spontaneous, 153–155
Genetic code, 173–174
Geocentric injection velocity, 277
Geocentricity, 8–10
  diagram, 2
Geometric albedo. *See* Albedo.
Germs, of life, 155
Gibbous phase, 43–44
Glass, iron, 216
Glucose
  labeled, 236
  starch, 193
Glutamic acid, 184
Glyceraldehyde, 188
Glycerol, cold, defense against, 213
Glycine, 162, 180
  labeled, 236
Goethite, 110
Golay detector, 133
Gram stain, 193
Gravity, of Mars, 62–63, 72
  equipotential surface, 90–91
Gravitation, universal constant of, 66
Gravitational acceleration. *See* Gravity, of Mars.
Green plants. *See* Plants, green.
Greenhouse effect, 122
  Mars, 211
Greenwich meridian, 19
Ground temperature. *See* Temperature, surface.
Guanidine, 187
Guanine, 165 ff.
  formation, 187
Gulliver, 236

H

Halophilic bacteria, 213, 218
Hard lander, 273
Haze, blue. *See* Haze, of Mars.
Haze, of Mars (*see also* Blue clearing; Clouds, of Mars), 59, 81, 138, 143–151
  albedo, 145
  blue clearing, 143, 146–147
  carbon black, 146–147
  carbon dioxide, crystals, 145
  characteristics, 143
  clearing, 143
  Earth's atmosphere, caused by, 150
  filters, study with, 143
  gases, atmospheric, 148–149
  interpretation, 144–148
  morning, 79
  polar, 139
  reflectivity, 145
  surface-contrast theory, 145
  ultraviolet radiation, 151

Heat
balance, 52
latent, of sublimation, 106
radiation, received by Earth, 52
Heliocentric, 10, 33
Helium
abundance, 74
carrier gas, 229
Helix, 164
deoxyribonucleic acid, 169
Hematite, 110
Hemispheres, of Mars
observation contrast, 48–49
seasonal contrast, 51, 54, 55
Heterotrophic organisms, 214
Hexamethylenetetramine, 188
High-energy protons. *See* Protons, high-energy.
High-temperature reactions, 183
Histone, 192, 193
Hoarfrost, 78, 106, 121
polar caps, 103–105
sublimation, 105
Hohmann trajectory, 274 ff.
Hohmann transfer ellipse, 274 ff.
Homolog, chemical, 199
Hot spring (*see also* Vulcanism), 188
Hourglass Sea, 16
Huygens' drawing, 16, 17
Hydration, water of, 110–111
Hydrocarbons
carbonaceous chondrites, 198–200
definition, 81
Martian atmosphere, 81
Hydrogen
abundance, 74
atmosphere, escape from, 83, 179
Hydrogen cyanide
formation, 180
ultraviolet radiation, 187
Hydrogen sulfide, in Martian atmosphere, 81
Hydrolysis, 187
Hydroxylamine
formaldehyde, reaction with, 183

I

Ice crystals, 78
Ice, in polar caps, 103–105
Ideal emitter. *See* Blackbody.
Ideal gas, 279
Ideal gas constant, 89
Ideal lapse rate, 95–96
Ideal radiator. *See* Blackbody.
Igneous rock, 110
Impact, accidental, spacecraft, 249, 252
avoidance, 270–271
Inclination
angle, 47
axis, 49–55
equatorial plane, 47
orbit, 55
Index of refraction, 121
Infrared measurements (*see also* Spectroscopy)
bright areas, 110
from spacecraft, 282–283, 295–296
polar caps, 107
temperatures, surface, 135
Injection energy, 275
Injection, trajectory, 276–278
Injection velocity, geocentric, 277
Insects, cold, defense against, 213
Insulin, 162, 164
molecular weight, 185
Integral albedo. *See* Albedo.
Interior, of Earth, 65, 67
Interior, of Mars, 61, 65, 66–68
differentiation, 68, 82
International Astronomical Union
map of Mars, 28
mass of Mars, 57
Ion
acceleration, 231
collection, 231
current, 230
definition, 92
Ion-exchange resins, 201
Ionization, cause, 92
Ionization chamber, 287
Ionosonde, 282
Ionosphere
Earth, 92
Mars, 92–94
night, effect, 93
regions, 92
Iron, 111
bright areas, 111
glass, 216
meteorite, 111, 196
oxides, 110
porphyrin, 201
Isomerism, optical, 164–165
Isostasy, 61–62

J

J-band, 234
life detector, 234

K

Kangaroo rat, 217
Kepler
  planetary motion, laws, 13–15
  planetary orbits, 12

L

Lactate, labeled, 236
Landers, 227
  observations, 283–285
  sterilization (*see also* Contamination of Mars, avoidance), 227
Landing
  hard, 273
  soft, 273
Landing capsule, 248
Lands, 21
Lapse rate, 95–96
Latent heat, of sublimation, 106
Latitude, 19
  duration of day, 51
Lava (*see also* Vulcanism), 119, 183
  microsphere production, 194
Leucine, 184
Levorotation
  amino acids, 166
  definition, 164–165
Lichens, 217
Life, definition, 158–159
Life, on Mars
  biochemistry, 210
  conditions for, 212–220
  cosmic rays, effect, 215–217
  detection, 225–246
    biological, 225–229
    chemical, 229–235
    metabolic, 236–239
    morphological, 240–242
    principles, 227–229
  development, 210–212
  evidence, 219–220
  factors affecting, 212–220
  Mariner IV, 220
  possibility, 209–223
  terrestrial organisms, survival, 218–219
  ultraviolet radiation, effect, 215–217
  water, importance, 217–218
Life, origin, 153–176
  archebiosis, 157
  Bible, Holy, 154
  historical review, 153–161
  membranes, 176
  molecular biology, 161–176
  nucleic acids, 173–175
  proteins, 176
  "soup," 158
Light
  polarization, 84–85
  scattering, 59
Lignite, 198
  fatty acids, 202
Limb
  brightening, 58–59
  darkening, 58–59
  definition, 44
Lime water, 188
Limonite, 109–111
  density, 111
  oxidizing agent, 213
  oxygen extraction, 222
  ultraviolet absorption, 216
Lineaments, 128, 292
Linear diameter, 58
Longitude, 19
Lowell's map, 25
Luciferase, 235
Luciferin, 235
Lysine, 184

M

Macromolecule, 164
Magma, 131
Magnetic field
  cosmic rays, effect, 215
  Mars, 64–65, 282
Magnetometer, 64, 282
Malic acid, 186
Mantle, Earth, 67
Maps of Mars, 18–31
  International Astronomical Union, 28
Mare (*see also* Dark areas), 111
Maria (*see also* Dark areas), 111
Mariner IV, 57, 210, 285–292
  attitude control, 285
  canals, 127–128
  craters, 128–130, 211
  description, 285–287
  instrumentation, 287
  ionosphere, 88, 93
  life on Mars, evidence, 220
  magnetic field measurement, 64
  magnetometer, 282
  occultation experiment, 87–97
  orientation, 286

Mariner IV—Continued
photographs, 280, 290–292
radius of Mars, determination, 60
television system, 288–290
trajectory, 87, 287–288
Mariner Mars 1969, 292–297
description, 292–293
instrumentation, 295–296
television system, 293–295
trajectory, 296–297
Mars
angle of inclination, 17
appearance, 15–31
atmosphere, 17, 73–97
atmospheric temperatures, 94–97
brightness, 41
characteristics, table, 72
clouds, 18, 137–143
colonization. *See* Colonization, of Mars.
color, 83
contamination, 218
cosmic radiation dose rate, 215
day, 63–64
day, length, 16, 64, 72
density, 62–63, 72
diameter, apparent, 41
distance to Sun, 55
drawing, earliest known, 15
Earth, flight time from, 276
elevation differences, 90
exobiology, 209–212
exploration, with spacecraft, 273
features, naming, 20–21
flattening, 72
gravity, 62–63, 72
habitation, requirements, 221–222
haze, 143–151
"inhabitants," 17, 29
interior, 65–68
ionosphere, 92–94
life, possibility (*see also* Life, on Mars), 209–223
magnetic field, 64–65, 282
maps, 18–31
mass, 72
moons. *See* Moons, of Mars.
observation (*see also* Observation, of Mars), 38–55
opportunity, 274
orbit, 13–15
orbital characteristics, 55
period of rotation, 16
phases, 15, 43–44
physical characteristics, 57–68
polar caps, 16–17
position in sky, 42–43, 47
prebiotic compounds, 210
radius, 72
rising, 42–43
rotation speed, 64
seasons (*see also* Seasons, on Mars), 52–55
setting, 42–43
soil, 244–245
spacecraft study. *See* Mariner.
speed of rotation, 72
Sun, distance from, 13
surface, 99–136
surface roughness, 227
symbol, 8
telescopic observation, 15–31
temperature, surface, 106, 131–136
temperatures, atmospheric, 94–97
volume, 60
water, availability, 211
water, precipitable, 80–81
Martian simulator, 217–218
Mass distribution, 61–62, 91
Mass, of Earth, 57
Mass, of Mars, 57, 72
determination, 57
Mass spectrometry, 231–232
carbonaceous chondrites, 199
gas chromatography, combination, 232–233
Mean solar day, 64
Melamine, 201
Membrane
and life origin, 174, 176
plasma (or cell), 160
Meridian, 63
prime, 19
Mesosphere, 95
Messenger ribonucleic acid, 172
Metaphosphate ester (*see also* Ethyl metaphosphate), 185
Meteor shower, 147–148
Meteorites (*see also* Carbonaceous chondrites), 68, 82
asteroidal belt, 111
carbonaceous chondrites, 196–206
chondrite, 111
definition, 196
elements, organized, 203–205
germs of life, origin, 155
iron, 111

stony, 111, 155, 196
types, 196–197
Methane, 74
atmosphere, of Mars, 81
detection, 226
in primitive atmosphere, 178
substituted, 81
Methionine, 164
Microcalorimetry, for life detection, 241
Microhills, 118–119
Micron, definition, 76
Micro-organism (*see also* Sterilization)
adaptability, 212
dieaway, 266
discovery, 154
in Martian environment, 218
meteorite contamination, 200
ultraviolet radiation, effect, 156
Microprobe, 205
Microscope, 154
vidicon, 241
Microspheres, 191–194
Gram stain, 193
Microwave, polarization, 227
Midcourse correction, 278, 288
Millibar, definition, 77
"Missing" planet, 205
Mitchel Mountains (or depressions), 102, 108
Mitosis, 160
deoxyribonucleic acid, 169
model, 193
Molecular biology, 161–176
Molecular-number density, 89, 91
Monomer, 164
Moon, of Earth
craters, 128
lineaments, 128
phases, 44
Moons, of Mars, 68–72
characteristics, 69–72
discovery, 21, 69
exploration, 72
perturbation of, 57
prediction, 68
Morning haze, 79
Morning star, 42
Morphology, life detection, 228
Motion
direct, 44–46
retrograde, 44–46
Mountains of Mitchel, 102, 108
mRNA, 172
Multiple evaluator. *See* Multivator.
Multivator, 237–239
Murray meteorite, 199 ff.
Mutation, 159, 161, 170–171
beneficial, 159
cosmic ray, 216
deleterious, 159

N

Naming, of features, 20–21
Neutrons, for surface analysis, 284
Nitrile, 181
Nitrogen
abundance, 74
atmosphere, of Mars, 89
origin, of atmospheric, 75
Nitrogen dioxide
atmosphere, of Mars, 81
blue clearing, 149
color, 149
polar caps, 109
quantity, in atmosphere, 109
yellow clouds, 143
Nitrogenous base, residues, 165
Noctilucent clouds, 148
Nomenclature, of features, 20–21
table, 26–27
Nucleic acids, 161, 165–173
configuration, 168
radiation damage, 211
synthesis, 174, 191
Nucleolus, 160
Nucleosides, 166
formation, 189–190
Nucleotides, 165
definition, 165
deoxynucleotide, 166
formation, 189–190
in nucleic acids, 165, 168
polymerization, 190–191
Nucleus, 160
Number density, atmosphere, 89, 91

O

"Oasis," 25
cloud location, 140
Oblateness. *See* Flattening, dynamical.
Obligate anaerobe, 178
Observation, of Mars, 38–55
visibility, 38

Occultation, 60
definition, 87
Mariner IV, 87–97, 287
Mariner 1969, 296
temperature, atmospheric, determination, 96
Ocean
formation, 83
life, origin, 211
Oligopeptide, 164
Opportunity, 274
Opposition (*see also* Aphelion; Conjunction; Perihelion), 37
aphelion, 39
axis, direction of, 49
blue clearing, 149–151
brightness, of Mars, 41
dates, 39–40
definition, 37
distance, Earth-Mars, 39–40, 55
favorable, 38–41, 46, 53–54
intermediate, 48
interval, 38, 40
perihelion, 39
phases, relation to, 43
unfavorable, 38–41, 46, 53–54
Optical activity
amino acids, 164–165
carbonaceous chondrites, 202–203
origin, 194–196
Pasteur probe, 233–234
substrate, 239
sugars, 166
Optical flattening (*see also* Flattening, dynamical), 60–61
Orbit
coasting, 277
determination, of Mars, 13
Earth, 14
eccentricity, 35
elliptical, 13, 34–36
inclination, 55
Mars, 13–15, 55
parking, 277
planetary, 13
Orbital characteristics, 55
Orbital plane, 42, 46, 47
Orbital speed, 4, 45, 55, 64
Orbiter spacecraft, 128, 226–227
observations, 282–283
Organic compounds, definition, 157
Orgueil meteorite, 199 ff.
Origin of life. *See* Life, origin.
Orthophosphoric acid, 185
Oxidation state, 187
Oxidizing agent, 213
Oxygen
abundance, 74
atmosphere, of Mars, 81–88
isotope exchange, for life detection, 238–239
isotopes, 238
origin, of atmospheric, 75
Ozone, in Earth's atmosphere, 95
Ozone, in Martian atmosphere, 81–82

P

Panspermia hypothesis, 155–156
Papain, 174, 185
Paraformaldehyde, 188
Parking orbit, 277
Partial pressure, 103
carbon dioxide, 108
Particles, charged, near Mars, 64
Pasteur probe, 233–234
Pentose, 165–166
Pepsin, 185
Peptide bond. *See* Peptide linkage.
Peptide linkage, 164
absorption band, 235
Peptone, 185
Perihelion (*see also* Aphelion; Conjunction; Opposition), 46
brightness, effect on, 41
definition, 36
distance, 36
duststorms, 141
Mars, 36
speed, 36
temperature, 132
Period
orbital, 55
rotational, 16
sidereal, 36–38
synodic, 36–38
Permafrost, 217–218
Mars, 82–83
Petroleum, 198
fatty acids, 202
origin, 199
Phase, Mars, 16, 43–44
Phase angle, 43–44
Phase contrast, microscope, 241
Phase integral. *See* Albedo.
Phase path, radio wave, 88–89

Phenylalanine, 163, 184
mass spectrum, 232
Phobos. *See* Moons, of Mars.
Phosphatase, determination for life detection, 238
Phosphorylation, 190
Photosynthesis, 75, 186, 214
Physical characteristics, of Mars, 57–68
table, 72
Phytane, 199–200
Phytol, 199
Plane polarization, 84–85
Planet, "missing," 205
Planet, name, derivation of, 7
Planetary motion, Kepler's laws (*see also* Planetary system), 13–15
Planetary quarantine, model, 257–258
Planetary system
Copernican, 10–11
Keplerian, 12–15
Ptolemaic, 9–10
Tychonic, 12
Planetoid. *See* Asteroid.
Planets, 7–15
discovery, 7–9
early observations, 7–31
motion, 7–15
orbits, 12–13, 33–38
speed, 14
Plants, green, 75
carbon dioxide utilization, 179, 214
chlorin, 201
chlorophyll, 200
photosynthesis, 214
Plasma, 92
Plasma diffusion, 92
Plasma membrane, 160
Plasma sheath, 280
Polar caps, 16–17, 18, 78, 99–109
albedo, 104
band around, 103, 105, 108
barometric pressure, 108
breakup, 101–102
carbon dioxide, 105–109
carbon dioxide, liquid, 108
carbon dioxide, tests for, 103–105
cloud cover, 99, 101
collar, 103–105, 108
composition, 103–105
dark band, 103
description, 99–100
elevation, 102, 105
frost, 103–105
hoarfrost, 103–105
ice, 78, 103–105
infrared measurements, 104, 107
location, 100
Mitchel Mountains (or depressions), 102, 108
nitrogen dioxide, 109
polarization, 104
recession, 100–103, 121
seasonal change, 78, 80, 99–100
snow, 103–105
spectroscopy, 104
sublimation, 103, 105
temperature, surface, 105–106
terrestrial, 99, 104
thickness, 104
visibility, through haze, 149
water, 103
wave of darkening, 117
Polar diameter, 59–60
Polarimeter, 202
Polarization, of light, 84–85
Polarization measurements
atmospheric, 85
clouds, of Mars, 137
polar cap, 104
surface, 109, 120–121
Polarized light
circularly, 195
plane, 85, 164–165
Polaroid, 85
Polymer, 164
Polymerase, DNA, 174
Polymerization, 188
amino acids, 164, 184–186
nucleotides, 190–191
Polynucleotides, formation, 190–191
Polypeptide, 163, 164
formation, 183–186
structural formula, 163
Polyphosphoric acid, condensing agent, 185–186
Porphyrin, 201–202
Position of Mars, in sky, 42–43, 46–49
Potassium-40, decay, 67, 78, 82
Prebiological (*or* prebiotic) materials, 177–206
definition, 177
production, 177–196
Precipitable water, 80–81
Pressure, atmospheric. *See* Pressure, barometric.

Pressure, barometric, 76, 83–87
carbon dioxide, solid, 121–122
determination, 83–87, 242
organisms, effect on, 215
polar caps, effect on, 108
polarimetry, 84–86
variation, 91–92
Pressure, partial. *See* Partial pressure.
Prime meridian, 19, 31
Pristane, 199–200
Probe, atmospheric, 273, 278–280
Proctor's map, 20
Proline, 163
Protein, 161–164
amino acids, 162
enzyme, 162, 173
precursor, 183
synthesis, by ribonucleic acid, 172–173
Proteinoid, 184–185
microspheres, 194
Proteolysis, 185
Proteolytic enzymes, 185
Protons, high-energy, 149, 244
life, on Mars, effect on, 215–217
Protozoa, 159
Ptolemaic planetary system, 9–10
Purines, 165
formation, 186–187
in nucleic acids, 165–173
pyrimidine, complementary, 169
Pyrimidines, 165
formation, 186–187
in nucleic acids, 165–173
purine, complementary, 169
Pyrolysis, 230

Q

Quadrature, 41–42, 79
phase angle, 44
Quarantine. *See* COSPAR specifications.
Quarantine, planetary, model, 257–258

R

Racemic mixture, 165
synthesis, laboratory, 195
Rad, 215
Radar reflection
canals, 127
dark areas, 123–124
observation, 90
Radiation belt, 64
Radiation, solar. *See* Solar radiation.
Radiator, ideal. *See* Blackbody.
Radioactive decay, 67, 78, 82
Radioisotope biochemical detector, 236–237
Radiometer, 226
Radiometric albedo. *See* Albedo.
Radius
Mars, 72
polar, 59–60
Radius vector, 14
Recession, of polar caps, 100–103
Reduced thickness, atmospheric, 77
Reducing agent, 213
Reducing atmosphere, 178, 183
electrical discharge, effect, 186
Reflection, random. *See* Scattering.
Reflection, specular, 83–84
Reflectivity (*see also* Albedo), 86, 112–113
Refraction, of radio waves, 88
Refractive index, 88–90, 121
Replication, 159
mechanism, 170–171
Resistance, aerodynamic, 278
Retrograde motion, 9, 14, 44–46
Ribonucleic acid, 168, 171–173
messenger, 172
protein formation, 172–173
radiation damage, 211
synthesis, 174
transfer, 172
Ribose, 165 ff.
Ribosome, 160
Rilles, 291
Rising, of Mars, 42–43
RNA. *See* Ribonucleic acid.
Rotation, optical (*see also* Dextrorotation; Levorotation), 164–165
Rotation speed, of Mars, 64, 72

S

Saltation, 119
Satellite. *See* Moon, of Earth; Moons, of Mars.
Satellite, artificial, 61–62
Satellite, of Mars. *See* Moons, of Mars.
Satellite, natural. *See* Moon, of Earth; Moons, of Mars.
Scale height, 89–90, 91, 96–97
Scattering, 59, 83
back, 146
forward, 145–146
Schiaparelli's nomenclature, 21
"Sea," 111

Seasons (*see also* Seasons, on Earth; Seasons, on Mars), 17
length, 51, 54
Seasons, on Earth, 50–52
opposition, 53–54
Seasons, on Mars, 52–55
dark areas, 112
eccentricity, effect, 54
length, 54
opposition, 53–54
polar caps, effect on, 99–100
Secular change
dark areas, 113, 116
Semipermeable membrane, 174
Serine, 184
Setting, of Mars, 42–43
Sidereal period, 36–38
Silica powder, chromatography, 229
Silicates, bright areas, 111
Simulator, Martian, 217–218
Sinton bands, 220
Sinus Meridianii, 17
Sky, position in, Mars, 42–43, 46–49
Slopes, surface, of Mars, 124
Snow
albedo, 104
craters, 128
polar caps, 103–105
Soft landing, 273
Soil
fatty acids, 202
Mars, 244–245
ultraviolet radiation protection, 216
Solar day, mean, 64
Solar-plasma probe, 287
Solar radiation, 66
absorption, 131–132
at Mars, 131–132
power supply, 223
Solar system, 33–38
Solar wind, 94
Solid particles, in Martian atmosphere, 86
Solstice, 51–53
"Soup," 158, 211
Spacecraft
fabrication, decontamination in, 266–270
observations from, 278–285
sterile assembly, 270
sterilization, 258–259
types, 273
Spectra, absorption, 73–74
Spectral absorption curve, 79
Spectrometry, mass. *See* Mass spectrometry.
Spectroscopy, 73–74, 76
infrared, 78, 80–81, 107, 110, 282–283, 295–296
life, on Mars, evidence, 219–220
polar caps, 104
Specular reflection, 83–84
Speed, orbital, 54–55
aphelion, 36
perihelion, 36
Speed, rotational, 64, 72
Spontaneous generation, 153–155
Spores
dormant state, 217
fungal, 216
Spot, flying, 241
Spring, 51
Spring equinox, 51
Martian, 53–55
Stain, Gram, 193
Standard atmosphere, 77
Stars, fixed, 7, 10
Stefan-Boltzmann equation, 131
Stereoisomerism (*see also* Optical activity), 233
Sterile assembly, of spacecraft, 270
Sterilization
gas, 259–260
heat, 262–266
importance, 247–248
liquid, 260–261
radiation, 261–262
solid chemical, 261
spacecraft, 247–271
Stony meteorites, 196
Stratosphere, 95
Strecker synthesis, 181, 183–184
Sugars
formation, 187–189
optical activity, 166
Summer, 51
Martian, 53–55
Summer solstice, 51
Sun-centered solar system. *See* Heliocentric.
Sun, distance from, 55
Sun sensor, 285
Surface-contrast theory, blue haze, 145, 150–151
Surface, of Mars (*see also* Bright areas; Dark areas), 99–136
particle size, variation, 122
polar caps, 99–109
silicates, 111
thermal conductivity, 133–134

Surface temperature. *See* Temperature, surface.
Swan bands, 280
Synodic period, 36–38
Synodic year, definition, 37
Synthesis, amino acids, 165
Syrtis Major, 16

T

Telescope, 11, 15
  inversion of image, 16
Television system
  of Mariner IV, 288–289
  of Mariner 1969, 292–294
  orbiter, 226
Temperature, ground. *See* Temperature, surface.
Temperature, of Martian atmosphere, 89, 94–97
Temperature, surface, 105–106, 131–136
  aphelion, 132
  bright areas, 121
  calculation, 131–132
  conclusions, 135
  dark areas, 121
  fluctuation, 130
  gradient, 107
  maximum, 135
  measurement, 132–135
  minimum, 135–136
  perihelion, 132
  solstice, 134, 135
Terminator, definition, 44
Testacidae, 216
Tests (shells), 216
Theoretical lapse rate, 95–96
Thermal conductivity, of Mars' surface, 133–134
Thermal creep, 130
Thermopile, 133, 295
Thermosphere, 95–96
Thickness, reduced, atmospheric, 77
Threonine, 184
Thymine, 165 ff.
Trajectory
  injection, 276–278
  interplanetary, 273–276
  type I, 275
  type II, 275
Transfer ellipse, 274
Transfer ribonucleic acid, 172
Transit, 72
Trapped-radiation detector, 287
Tropopause, 96
Troposphere, 95–96
tRNA, 172
Trypsin, 185
Tryptophan, 184
"Twinning," canals, 24–25, 29
Tychonic planetary system, 12

U

Ultraviolet radiation, 182 ff., 216
  atmosphere, primitive, effect on, 178 ff.
  bacteriocidal property, 156
  blue haze, 151, 216
  carbon dioxide decomposition, 81
  fungi, effect on, 216
  hydrogen cyanide, 187
  ionization, 92, 149
  life on Mars, effect on, 215–217, 222
  soil, absorption by, 216
  water decomposition, 111
Universal gas constant, 89
Uracil, 165 ff.
  cytosine, conversion from, 186
  polynucleotide, synthetic, 191
  prebiotic synthesis, 186
Urea, 186

V

Vacuum thermopile, 133
Valine, 173
Vanadyl porphyrin, in meteorite, 201–202
Van Allen belt, 64
Vegetation, on Mars
  dark areas, 111
  evidence, 219
  wave of darkening, 117–118
Vernal equinox, 51
Vidicon camera, 293
Vidicon microscope, 241
Violet layer. *See* Haze, of Mars.
Virus replication, 175
Volcano. *See* Vulcanism.
Volume of Mars, 60
Vulcanism, 75, 82–83, 217
  craters, orgin, 130–131
  energy source, 179
  high-temperature reactions, 183

W

Water (*see also* Water vapor)
  of hydration, 110, 120–121
  on Mars, 131, 211, 217
  polar caps, 103
  precipitable, 80–81
  soil, 118
  ultraviolet decomposition, 111

Water-cell transmission, 133
Water vapor, 74
  Martian atmosphere, 78–81
  mass, in atmosphere, 105
Wave of darkening, 114–120
  canals, 127
  duststorms, 118
  equator, 115
  explanation, 117–120
  life, evidence, 219
  microhills, 118–119
  photometry, 115
  polar cap recession, 117
  vegetation, 117–118
W-cloud group, 140
Western quadrature, 41–42
White clouds, 138–141
Wind, 243
  micro-organisms, distribution, 243
  power supply, 223
  solar, 94
  velocity, 127, 141
Winter, 51
  Martian, 53–55
Winter solstice, 51
Wolf trap, 240
Wright effect, 58–59
Wright's phenomenon. *See* Wright effect.

X

X-rays, ionization by, 149

Y

Year
  sidereal, 36–38
  synodic, 36–38
Yellow clouds, 81, 141–143

Z

Zenith, definition, 48